Advances in Spectral-Spatial Classification of Hyperspectral Images

Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, James Tilton

INVITED PAPER

Advances in Spectral–Spatial Classification of Hyperspectral Images

By Mathieu Fauvel, Yuliya Tarabalka, *Member IEEE*, Jón Atli Benediktsson, *Fellow IEEE*, Jocelyn Chanussot, *Fellow IEEE*, and James C. Tilton, *Senior Member IEEE*

ABSTRACT | Recent advances in spectral–spatial classification of hyperspectral images are presented in this paper. Several techniques are investigated for combining both spatial and spectral information. Spatial information is extracted at the object (set of pixels) level rather than at the conventional pixel level. Mathematical morphology is first used to derive the morphological profile of the image, which includes characteristics about the size, orientation, and contrast of the spatial structures present in the image. Then, the morphological neighborhood is defined and used to derive additional features for classification. Classification is performed with support vector machines (SVMs) using the available spectral information and the extracted spatial information. Spatial postprocessing is next investigated to build more homogeneous and spatially consistent thematic maps. To that end, three presegmentation techniques are applied to define regions that are used to regularize the preliminary pixel-wise thematic map. Finally, a multiple-classifier (MC) system is defined to produce relevant markers that are exploited to segment the hyperspectral image with the minimum spanning forest algorithm. Experimental results conducted on three real hyperspectral images with different spatial and spectral resolutions and corresponding to various contexts are presented. They highlight the importance of spectral–spatial strategies for the accurate classification of hyperspectral images and validate the proposed methods.

KEYWORDS | Classification; hyperspectral image; kernel methods; mathematical morphology; morphological neighborhood; segmentation; spectral–spatial classifier

I. INTRODUCTION

Recent advances in hyperspectral remote sensor technology allow the simultaneous acquisition of hundreds of spectral wavelengths for each image pixel. This detailed spectral information increases the possibility of more accurately discriminating materials of interest. Further, the fine spatial resolution of the sensors enables the analysis of small spatial structures in the image. Many operational imaging systems (Table 1) are currently available providing a large amount of images for various thematic applications.

- Ecological science. Hyperspectral images are used to estimate biomass, biodiversity, or to study land cover changes [1]–[3].
- Geological science. It is possible to recover physicochemical mineral properties such as composition and abundance [4].
- Hydrological science. Hyperspectral imagery is used to determine changes in wetland characteristics [5]. Water quality, estuarine environments, and coastal zones can be analyzed as well.
- Precision agriculture. Hyperspectral data are used to classify agricultural classes and to extract nitrogen content for the purpose of precision agriculture [6], [7].

Manuscript received November 1, 2011; revised February 2, 2012; accepted April 12, 2012.

M. Fauvel is with the DYNAFOR lab, University of Toulouse-INRA, Castanet-Tolosan 31326, France (e-mail: mathieu.fauvel@ensat.fr).
Y. Tarabalka is with AYIN, INRIA, Sophia Antipolis F-06902, France.
J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik IS 107, Iceland.
J. Chanussot is with the GIPSA-lab, Grenoble Institute of Technology, Grenoble 38000, France.
J. C. Tilton is with the NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA.

Digital Object Identifier: 10.1109/JPROC.2012.2197589

Table 1 Examples of Operational Systems

Instrument	Spectral range (nm)	Bands	Bandwidth (nm)
AVIRIS	400-2500	224	10
HYDICE	400-2500	210	10
ROSIS-03	430-860	115	4
HyMAP	400-2500	126	10-20
CASI	400-870	288	1.9
HYPERION	400-2500	200	10

- Military applications. The rich spectral spatial information can be used for target detection [8], [9].

The intrinsic properties of hyperspectral images need to be addressed specifically because conventional classification algorithms made for multispectral images do no adapt well to the analysis of hyperspectral images [10]. Two major challenges have been identified this last decade: the *spectral dimensionality* and the need for specific *spectral–spatial classifiers*.[1]

In the spectral domain, pixels are represented by vectors for which each component is a measurement corresponding to specific wavelengths [11]. The size of the vector is equal to the number of spectral bands that the sensor collects. For hyperspectral images, several hundreds of spectral bands of the same scene are typically available, while for multispectral images, up to ten bands are usually provided. With increasing dimensionality of the images in the spectral domain, theoretical and practical problems arise. The idea of the *dimension* is intuitive, driven by experiments in 1-D, 2-D, or 3-D spaces, and geometric concepts that are self-evident in these spaces do not necessarily apply in higher dimensional spaces [12], [13]. For example, in high-dimensional spaces, normally distributed data have a tendency to concentrate in the tails, which seems to be contradictory with its bell-shaped density function [14]. Moreover, the rate of convergence of the statistical estimation decreases when the dimension grows while conjointly the number of parameters to estimate increases, making the estimation of the model parameters very difficult [15]. Consequently, with a limited training set, beyond a certain limit, the classification accuracy actually decreases as the number of features increases [16]. For the purpose of classification, these problems are related to the *curse of dimensionality*.

Intensive work has been performed in the remote sensing community in the last decade to build accurate classifiers for hyperspectral images. Bayesian models [12], feature extraction and feature reduction techniques [12], [17], random forest [18], neural networks [19], and kernel methods [20] have been investigated for the classification of such images. In particular, support vector machines (SVMs) have shown remarkable performance in terms of classification accuracy when a limited number of training samples is available [21]. SVMs perform a nonlinear pixel-wise classification based on the full spectral information which is robust to the spectral dimension of hyperspectral images [22]. Yet, the SVMs (and other pixel-wise methods) classify the image without using contextual information, i.e., the interpixel dependency. Hence, the hyperspectral image is treated as a list of spectral measurements with no spatial organization [23].

A joint spectral classifier is needed to reduce the labeling uncertainty that exits when only spectral information is used, helping to overcome the salt-and-pepper appearance of the classification. Further, other relevant information can be extracted from the spatial domain: for a given pixel, it is possible to extract the size and the shape of the structure to which it belongs. This information will not be the same if the pixel belongs to a roof or to a green area. This is also a way to discriminate between various structures made of the same materials. If spectral information alone is used, the roofs of a private house and of a larger building will be detected as the same type of structure. But using additional spatial information—the size of the roof, for instance—it is possible to classify these into two separate classes [24].[2]

Landgrebe and Kettig were probably the first to propose a classifier that used contextual and spectral information, the well-known ECHO classifier [12], [26]. Later, Landgrebe and Jackson proposed an iterative statistical classifier based on Markov random field (MRF) modeling [27], [28]. MRF modeling has been shown to perform well for the classification of remote sensing images [29], [30]. However, classical MRF modeling (e.g., Ising, Potts) suffers from the high spatial resolution: neighboring pixels are highly correlated, and the standard neighbor system definition does not contain enough samples to be effective. Unfortunately, a larger neighbor system imposes intractable computational problems, thereby limiting the benefits of conventional MRF modeling. Furthermore, algorithms involving MRF-based strategies traditionally require an iterative optimization step, such as simulated annealing, which is extremely time consuming. Recent works on graph-cut methods have reduced the processing time [31], [32]. Actually, these methods have only been applied to images with few spectral components, such as SAR images [33]. However, they are promising tools. Note that recently adaptive MRF have been introduced in remote sensing [34], [35] and, as graph-cut methods, are promising techniques.

Using the same crisp neighbor set employed by MRFs, textural features can be also extracted from the image [36]. Texture features have been widely used in remote sensing; see, for instance, [37] and [38]. They provide relevant information about the granularity of the surface. However,

[1]Multispectral images need a spectral–spatial classifier as well. But the complexity makes the conventional spectral–spatial classifiers perform badly on hyperspectral image.

[2]Classification is only discussed in this paper, but other processings take benefit of combining spatial and spectral information, e.g., in unmixing application [25].

the texture features (entropy, variance, etc.) are usually computed in a moving window, thus imposing a crisp and common neighbor set for every pixel in the image.

Benediktsson *et al.* have proposed to use advanced morphological filters as an alternative way of performing joint classification [39]. Rather than defining a crisp neighbor set for every pixel, morphological filters enable the adaptive definition of the neighborhood of a pixel according to the structures to which it belongs to. Adaptive neighborhood approaches have given good results for multispectral and hyperspectral data [40]–[42]. More generally, the authors have previously used morphological processing to analyze the interpixel dependency at the object level. SVM and kernel functions were used to combine the spatial and spectral information during the classification process.

Another approach for including spatial information in the classification process starts with the performance of image segmentation. Segmentation methods partition an image into nonoverlapping homogeneous regions with respect to some criterion of interest, or homogeneity criterion (e.g., based on the intensity or on the texture) [43]. Hence, each region in the segmentation map defines a spatial neighborhood for all the pixels within this region. This approach extracts large neighborhoods for large homogeneous regions, while not missing small regions consisting of one or a few pixels. Different techniques have been investigated for hyperspectral image segmentation, such as watershed, partitional clustering, and hierarchical segmentation (HSeg) [44]–[47]. Then, the SVM classifier and majority voting are applied for combining spectral and spatial information: for every region in a segmentation map, all the pixels are assigned to the most frequent class within this region, based on SVM classification results [45]. The described approach leads to an improvement of classification accuracies when compared with spectral–spatial techniques using local neighborhoods for analyzing spatial information.

However, automatic segmentation of hyperspectral images is a challenging task, because its performance depends both on the chosen measure of region homogeneity and on the parameters involved in the algorithm. An alternative way to get accurate segmentation results consists in applying a marker-controlled segmentation [43], [48]. The idea is to select for every spatial object one or several pixels belonging to this object, called a *marker*, or a seed of the corresponding region. Then, regions are grown from the selected seeds, resulting in a segmentation map. The region markers can be chosen either manually, which is time consuming, or automatically. In the automatic approach, a probabilistic classification is applied to the data, and then the most reliably classified pixels, i.e., pixels belonging with the high probability to the assigned class, are selected as markers of spatial regions [46], [49]. The decision about which pixels to retain as markers is based on the results of either a single probabilistic SVM classifier, or a multiple-classifier (MC) system. Furthermore, a marker-controlled segmentation algorithm can be applied by building a minimum spanning forest (MSF) algorithm rooted on the selected seeds. By assigning the class of each marker to all the pixels of the region grown from this maker, a spectral–spatial classification map is obtained.

The main objective of this paper is to present recent advances in techniques for the classification of hyperspectral images, which face the following issues:

- the limited training samples;
- the extraction of spatial features;
- the spectral–spatial classification of the image.

The remainder of this paper is organized as follows. Section II presents three hyperspectral images with high spatial resolution that will be used for experiments throughout the paper. Section III provides a general framework for the classification of remote sensing hyperspectral images. Section IV focuses on the spectral–spatial classification with morphological features. Basics of mathematical morphology are reviewed, then several concepts (morphological profile, morphological neighborhood) are presented with classification methods that include spatial features in the process. Section V explores classification using segmentation-derived adaptive neighborhoods. Three different segmentation techniques are presented, then a spectral–spatial classification scheme combining segmentation and pixel-wise classification maps is described. Section VI discusses segmentation and classification of hyperspectral images using automatically selected markers. Finally, conclusion and perspectives are given in Section VII. Table 2 summarizes the notations used in this paper.

II. DATA SETS

Three high spatial resolution hyperspectral data sets are used in this paper. Two images of an urban area were acquired with the Reflective Optics System Imaging Spectrometer (ROSIS-03) optical sensor. The flight over the city of Pavia, Italy, was operated by the Deutschen Zentrum für Luft- und Raumfahrt (DLR, German Aerospace Agency) within the context of the HySens project, managed and sponsored by the European Union. According to specifications, the ROSIS-03 sensor provides 115 bands with a spectral coverage ranging from 0.43 to 0.86 μm. The spatial resolution is 1.3 m per pixel. The two data sets are as follows.

1) *University Area*: The first test set took place near the Engineering School, University of Pavia, Pavia, Italy. It was 610 $\times$ 340 pixels. Twelve channels were removed due to noise. The remaining 103 spectral channels were processed. Nine classes of interest were considered: tree, asphalt, bitumen, gravel, metal sheet, shadow, bricks, meadow, and soil.
2) *Pavia Center*: The second test set was the center of Pavia. The image was originally 1096 $\times$ 1096 pixels. A 381-pixel-wide black band in the left-hand side part of image was removed, resulting in a "two-part" image of 1096 $\times$ 715 pixels.

Table 2 Notations and Acronyms

$\mathbf{x}$	A pixel-vector ($\mathbf{x} \in \mathbb{R}^d$)
d	Number of spectral wavelengths
y	Class label of a given pixel
I	An image
SAR	Synthetic-Aperture Radar
ROSIS	Reflective Optics System Imaging Spectrometer
AVIRIS	Airborne Visible/InfraRed Imaging Spectrometer
MM	Mathematical Morphology
B or SE	Structuring Element
ϵ_B	Erosion with B
δ_B	Dilation with B
γ_B (γ_B^r)	Opening (Opening by reconstruction)
ϕ_B (ϕ_B^r)	Closing (Closing by reconstruction)
MP	Morphological Profile
EMP	Extended MP
PCA	Principal Component Analysis
KPCA	Kernel PCA
NWFE	Nonparametric Weighted Feature Extraction
DBFE	Decision Boundary Feature Extraction
BDFS	Bhattacharyya Distance Feature Selection
ASF	Alternate Sequential Filter
$\Omega_{\mathbf{x}}$	Morphological neighborhood of a pixel $\mathbf{x}$
Φ	Mapping matrix
k (k_σ)	Kernel function (Gaussian kernel)
SVM	Support Vector Machines
MRF	Markov Random Field
K	Number of classes
MKL	Multiple Kernel Learning
RCMG	Robust Color Morphological Gradient
$\mathcal{X}$	Set of pixels inside a SE
EM	Expectation Maximization algorithm
CEM	Classification Expectation Maximization
C	Number of clusters
$\boldsymbol{\mu}_c$	Mean vector of cluster c
$\boldsymbol{\Sigma}_c$	Covariance matrix of cluster c
ω_c	Proportion of cluster c
PCFA	Piecewise Constant Function Approximations
HSWO	Hierarchical StepWise Optimization
MC	Multiple Classifiers
SAM	Spectral Angle Mapper
MSF	Minimum Spanning Forest
MSSC	Multiple Spectral-Spatial Classifier
WH	Watershed
MV	Majority Voting
G	Undirected graph
(V,E,W)	Set of vertices, edges and weights of the graph
F	Spanning forest
Z	McNemar's test

Thirteen channels have been removed due to noise. The remaining 102 spectral channels were processed. Nine classes of interest were considered: water, tree, meadow, brick, soil, asphalt, bitumen, tile, and shadow.

Available training and test sets for each data set are given in Tables 3 and 4, respectively. These are pixels selected from the data by an expert, corresponding to predefined species/classes. Pixels from the training set are excluded from the test set in each case and *vice versa*. Figs. 1 and 2 present false color images of the two ROSIS-03 data sets.

The third hyperspectral image was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the agricultural Indian Pine test site in Northwestern Indiana. The image has spatial dimensions of 145 × 145 pixels with a spatial resolution of 20 m per pixel. Twenty water absorption bands (104–108, 150–163, 220)

Table 3 Information Classes and Training-Test Samples for the University Area Data Set

Class		Samples	
No	Name	Train	Test
1	Asphalt	548	6304
2	Meadow	540	18146
3	Gravel	392	1815
4	Tree	524	2912
5	Metal Sheet	265	1113
6	Bare Soil	532	4572
7	Bitumen	375	981
8	Brick	514	3364
9	Shadow	231	795
Total		3921	40002

Table 4 Information Classes and Training-Test Samples for the Pavia Center Data Set

Class		Samples	
No	Name	Train	Test
1	Water	824	65971
2	Tree	820	7598
3	Meadow	824	3090
4	Brick	808	2685
5	Bare Soil	820	6584
6	Asphalt	816	9248
7	Bitumen	808	7287
8	Tile	1260	42826
9	Shadow	476	2863
Total		7456	148152

were removed [50], and a 200-band image was used for the experiments. The reference data contain 16 classes of interest, which represent mostly different types of crops and are detailed in Table 5. A three-band false color image and the reference data are presented in Fig. 3. In Fig. 3, 50 samples for each class were randomly chosen from the reference data as training samples, except for classes "*alfalfa*," "*grass/pasture mowed*," and "*oats*." These classes

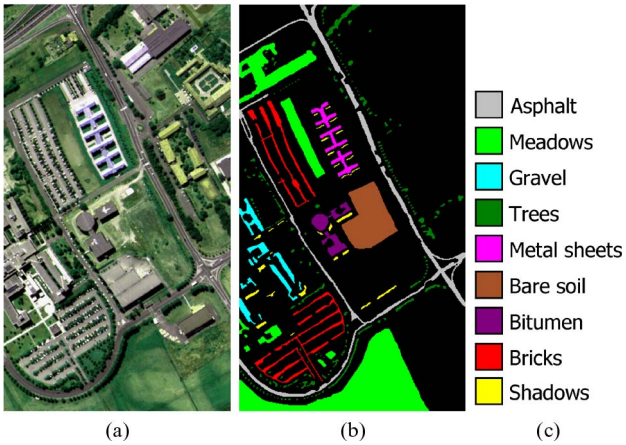

Fig. 1. ***University Area image. (a) Three band false color composite. (b) Reference data. (c) Color code.***

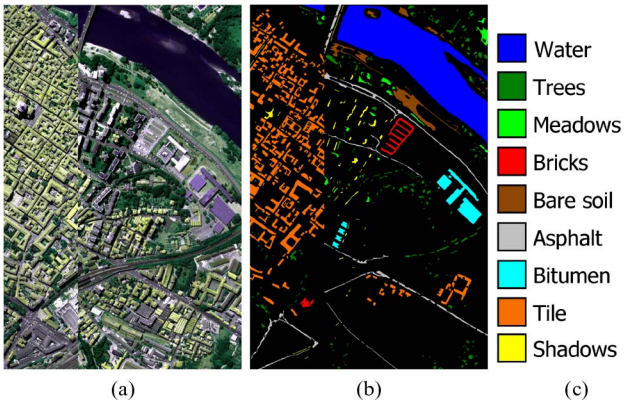

Fig. 2. ***Pavia Center image. (a) Three band false color composite. (b) Reference data. (c) Color code.***

contain a small number of samples in the reference data. Therefore, only 15 samples for each of these classes were chosen randomly to be used as training samples. The remaining samples composed the test set.

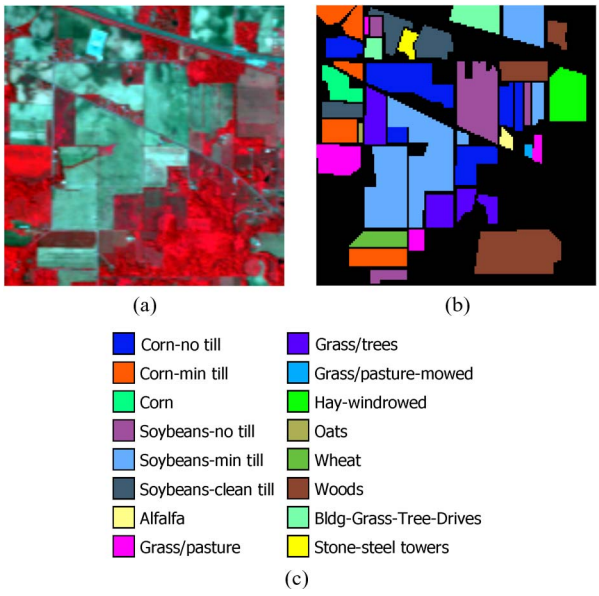

Fig. 3. ***Indian Pines image. (a) Three-band color composite. (b) Reference data. (c) Color code.***

III. GENERAL FRAMEWORK FOR THE CLASSIFICATION OF REMOTE SENSING HYPERSPECTRAL IMAGES

A general framework typically used for the classification of hyperspectral images is given in Fig. 4. The gray portion represents the area of research covered by the paper. The first step consists of extracting meaningful information from the data. It is done in the spectral domain [principal component analysis (PCA), decision boundary feature extraction (DBFE), nonparametric weighted feature extraction (NWFE), and kernel PCA (KPCA)] and in the spatial domain (mathematical morphological and hyperspectral segmentation). In extracting features in the spatial domain, the original contribution of this work is that the analysis is done at the object level and not a the pixel level. Hence, the approaches are adaptive in the sense that the local neighborhood of a pixel is taken into account when extracting the spatial information. The proposed methods are explained in Sections IV-A–IV-C and V-A–V-C.

The second original contribution of the work concerns the strategies developed to combine the spectral and spatial features that have been extracted. Several strategies are proposed: feature fusion (Section IV-D1), composite kernel (Section IV-D2), and spatial regularization by majority voting (Section V-D) or MSF (Section VI).

Finally, spatial regularization is investigated to postprocess the classification map. Several strategies are proposed. The first one, majority voting, uses a presegmentation map; see Section V. The second one is based on the MSF; see Section VI.

Table 5 Information Classes and Number of Labeled Samples for the Indian Pines Data Set

Class		Samples	
No	Name	Train	Test
1	Corn-no till	50	1384
2	Corn-min till	50	784
3	Corn	50	184
4	Soybeans-no till	50	918
5	Soybeans-min till	50	2418
6	Soybeans-clean till	50	564
7	Alfalfa	15	39
8	Grass/pasture	50	447
9	Grass/trees	50	697
10	Grass/pasture-mowed	15	11
11	Hay-windrowed	50	439
12	Oats	15	5
13	Wheat	50	162
14	Woods	50	1244
15	Bldg-Grass-Tree-Drives	50	330
16	Stone-steel towers	50	45
Total		695	9671

IV. CLASSIFYING HYPERSPECTRAL IMAGES WITH SPATIAL FEATURES EXTRACTED WITH MATHEMATICAL MORPHOLOGY

Mathematical morphology (MM) is a theory for nonlinear image processing [51], [52]. Morphological operators have already proven their potential in remote sensing image processing [53]. Several techniques have been considered with MM, ranging from image segmentation to automatic extraction of objects of interest [53], [54]. In the following, morphological operators are reviewed. Attention is paid to MM tools that allow the analysis of the image at the region level for the purpose of classification. Then, the concepts

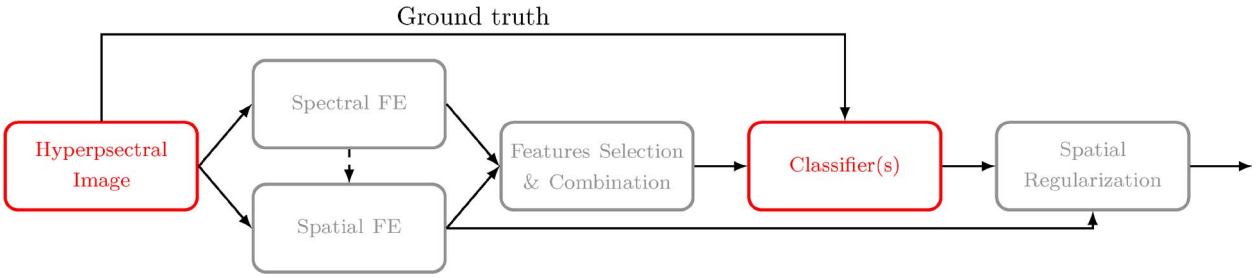

Fig. 4. ***General framework for the classification of hyperspectral images. FE means feature extraction.***

of morphological profile and morphological neighborhood are presented.

A. Morphological Operators

MM aims to analyze spatial relationships between pixels using a set of known shape and size (e.g., disk of radius 3 pixels), called the structuring element (SE) [48]. The two basic MM operators are *erosion* and *dilation*. Consider an image I and the value of the image for a given pixel $\mathbf{x}$, $I(\mathbf{x}) \in \mathbb{R}$. The result of an erosion $\epsilon_B(I(\mathbf{x}))$ of an image I at a pixel $\mathbf{x}$ by a structuring element B is the *minimum* value of pixels inside $B_\mathbf{x}$ ($B_\mathbf{x}$ is B centered at pixel $\mathbf{x}$)

$$\epsilon_B(I(\mathbf{x})) = \min_{\mathbf{x}_i}(I(\mathbf{x}_i) \in B_\mathbf{x}). \tag{1}$$

The dilation δ is defined as the dual operator, and the min operator is switched to the max operator

$$\delta_B(I(\mathbf{x})) = \max_{\mathbf{x}_i}(I(\mathbf{x}_i) \in B_\mathbf{x}). \tag{2}$$

The erosion expands objects of the image that are darker than their surrounding, while the dilation shrinks them (and *vice versa* for objects that are brighter than their surrounding). Moreover, bright (respectively, dark) structures that cannot contain the SE are removed by erosion (dilation). Hence, both erosion and dilation are noninvertible transformations.

Combining erosion and dilation, *opening* and *closing* operators can be defined. The opening $\gamma_B(I)$ is defined as the erosion of I by B followed by the dilation with B[3]

$$\gamma_B(I) = \delta_B \circ \epsilon_B(I). \tag{3}$$

The idea to dilate the eroded image is to recover most structures of the original image, i.e., structures that were not removed by the erosion and are bigger than B. The closing $\phi_B(I)$ is defined as the dilation of I by B followed by the erosion with B

$$\phi_B(I) = \epsilon_B \circ \delta_B(I). \tag{4}$$

Hence, with opening or closing, it is possible to get, for a given size of B, which structures (buildings, roads, etc.) of the image are smaller than B. However, opening and closing operators are not connected filters. For instance, two buildings can be merged into one, and thus, for instance, bias the analysis of the size distribution; see Fig. 5.

In order to avoid that problem, connected operators such as geodesic operators can be used [55]. The *geodesic dilation* $\delta_J^{(1)}(I)$ of size 1 consists in dilating an image (marker) I with respect to a mask J

$$\delta_J^{(1)}(I) = \min\left(\delta^{(1)}(I), J\right). \tag{5}$$

In general, I is the eroded image of J. Similarly, the *geodesic erosion* $\epsilon_J^{(1)}(I)$ is the dual transformation of the geodesic dilation

$$\epsilon_J^{(1)}(I) = \max\left(\epsilon^{(1)}(I), J\right) \tag{6}$$

and in that case I is the dilated image of J. The geodesic dilation (erosion) of size n is obtained by performing n successive geodesic dilations (erosions) of size 1 and leads to the definition of reconstruction operators. The *reconstruction by dilation* (*erosion*) of a marker I with respect to a mask J consists in repeating a geodesic dilation (erosion) of size 1 until stability, i.e.,

$$\delta_J^{(n+1)}(I) = \delta_J^{(n)}(I)\left(\epsilon_J^{(n+1)}(I) = \epsilon_J^{(n)}(I)\right). \tag{7}$$

[3]In this work, only symmetric SEs are considered. Otherwise, the symmetrical representation of B must be used in the opening/closing [48].

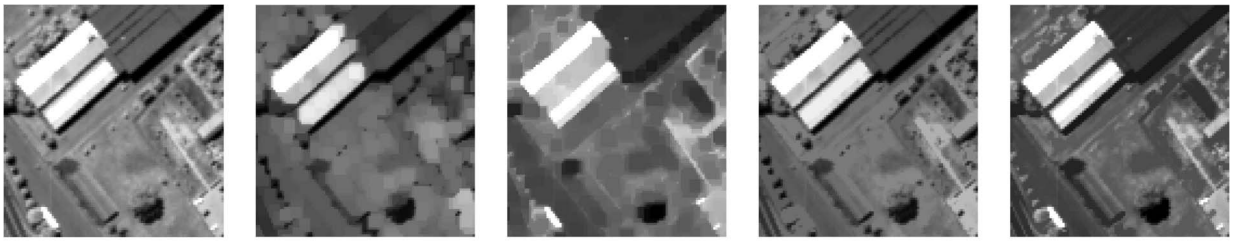

Fig. 5. *(a) Original image. (b) Opened image. (c) Closed image. (d) Geodesicaly opened image. (e) Geodesicaly closed image. The SE was a disk of radius 3 pixels. It can be seen in (c) that with the conventional closing the two bright buildings are merged into one. This is not the case with the geodesic operator.*

With these definitions, it is possible to define connected transformations that satisfy the following assertion: if the structure of the image cannot contain the SE then it is totally removed; else it is totally preserved. These operators are called *opening/closing by reconstruction* [55]. The *opening by reconstruction* $\gamma_r^{(n)}(I)$ of an image I is defined as the reconstruction by dilation of I from the erosion with an SE of size n of I. *Closing by reconstruction* $\phi_r^{(n)}(I)$ is defined by duality. Examples of opening/closing by reconstruction are given in Fig. 5.

B. Morphological Profile

Using opening/closing by reconstruction, it is possible to determine the size of the different structures of the image [56]. For a given size of the SE, it is possible to get structures which are smaller (they are removed) or bigger (they are preserved) than the SE. Applying such operators with a range of SE of growing size, one can extract information about the contrast and the size of the structures present in the image. This concept is called *granulometry*. The morphological profile (MP) of size n has been defined as the composition of a granulometry of size n built with opening by reconstruction and a (anti)granulometry of size n built with closing by reconstruction

$$\mathrm{MP}^{(n)}(I) = \left[\phi_r^{(n)}(I), \ldots, \phi_r^{(1)}(I), I, \gamma_r^{(1)}(I), \ldots, \gamma_r^{(n)}(I)\right]. \tag{8}$$

From a single panchromatic image, the MP results in a $(2n+1)$-band image. An example of MP is given Fig. 6. Its use for the classification of panchromatic images has shown a good improvement in terms of classification accuracy [39], [57]–[59]. However, when considering multivalued images, such as hyperspectral images, the direct construction of the MP is not straightforward, because of the lack of ordering relation between vector. In order to overcome this shortcoming, several approaches have been considered (see [60] for a review of several multivariate morphological filters). Our method, namely, the extended morphological profile (EMP),[4] consists in extracting a few images from the hyperspectral data that contain most of the spectral information by some dimension reduction method. The EMP was first proposed with PCA [40], [62], but it was also computed with independent component analysis (ICA) [63], KPCA [64], NWFE, DBFE, and Bhattacharyya distance feature selection (BDFS) [65]. Consider the m first principal components extracted from the hyperspectral image with PCA. For all components, the MPs are built. Then, they are stacked to construct the EMP

$$\mathrm{EMP}_m^{(n)}(I) = \left[\mathrm{MP}_1^{(n)}(I), \ldots, \mathrm{MP}_m^{(n)}(I)\right]. \tag{9}$$

The EMP contains some of the original spectral information, selected with some feature extraction algorithms, and some spatial information extracted with the morphological operators. The EMP can be used as an input to the classifier, or it can be fused with other information. The different strategies are discussed in Section IV-D.

C. Morphological Neighborhood

Geodesic opening/closing operators are appropriate in remote sensing because they preserve shapes. However, they cannot provide a complete analysis of remotely sensed images because they only act on the extrema (clear or dark objects) of the image [66], [67]. Moreover, some structures may be darker than their neighbors in some parts of the image, yet lighter than their neighbors in others, depending on the illumination. Although this problem can be partially addressed by using an alternate sequential filter (ASF) [68], the MP thus provides an incomplete description of size structures distribution. Fig. 7 illustrates this phenomenon.

[4]Note that this definition is somewhat different from [61].

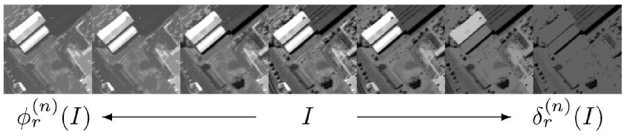

Fig. 6. *Morphological profile constructed with three opening/closing by reconstruction with a circular SE of size 2, 6, and 10. The left-hand side part corresponds to the closings by reconstruction and the dark objects are progressively deleted, e.g., the shadow of the big tree in the middle of the image. The right-hand side part corresponds to the openings by reconstruction and the bright objects are progressively deleted, e.g., the buildings in the upper part of the image.*

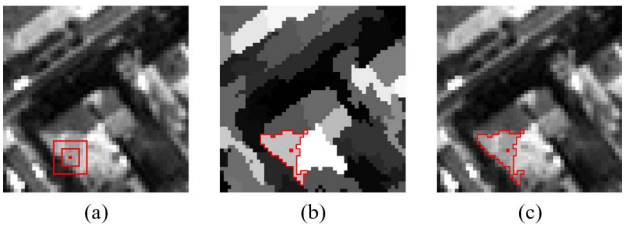

Fig. 8. *Morphological neighborhood. (a) Original image and fixed square neighborhood (in red). (b) Filtered image and neighbor set defined using area flat zones filter of size parameter $\lambda = 30$ [66]. (c) Original image with the defined neighbor set $_{\mathbf{x}}$. Illustration taken from [67].*

Another approach consists in defining an adaptive neighbor system for each pixel, the *morphological neighborhood* Ω. The morphological neighborhood of a pixel $\mathbf{x}$, $\Omega_{\mathbf{x}}$, is defined as the set of pixels that belongs to the same spatial structure as $\mathbf{x}$. This concept is connected to the more general concept of adaptive image neighborhood in image processing [69], [70]. Our approach developed in [67] uses a *self-complementary area filter* [66] to extract consistent spatially connected components. A self-complementary area filter is a filter that removes all structures of the image smaller (in terms of number of pixels) than a user-defined threshold; see Figs. 7 and 8. The filtered image is partitioned into flat zones. Each flat zone belongs to one single structure in the original image, as can be seen in Fig. 8(b). Furthermore, the smallest structures are removed and only the main structures of interest remain. The morphological neighborhood $\Omega_{\mathbf{x}}$ was defined as the set of pixels that belong to the same flat zone in the filtered image. The neighborhoods defined in this way are applied to the original image. This neighborhood is obviously more homogeneous and spectrally consistent than the conventional eight-connected fixed square neighborhood; see Fig. 8.

Similar to the MP, applying this filter on hyperspectral images is not possible because of the lack of an ordering relation. The same strategy is proposed, which consists in extracting one principal component from which the mophological neighborhood is computed. Then, the neighborhood mask is applied on each band of the data. Once the neighborhood of each pixel is adaptively defined, the spatial information is extracted: the vector median value of the neighbors set $\Omega_{\mathbf{x}}$ is computed for every pixel $\mathbf{x}$ [71]

$$\Upsilon_{\mathbf{x}} = \mathrm{med}(\Omega_{\mathbf{x}}) \tag{10}$$

where $\dim(\mathbf{x}) = \dim(\Upsilon_{\mathbf{x}}) = d$, the number of spectral bands. Unlike the mean vector, the median vector is a vector from the initial set, which ensures a certain spectral consistency since no new spectral values are created.

In conclusion, by defining the morphological neighborhood, every pixel has two features: the *spectral feature* $\mathbf{x}$, which is the original value of each pixel, and the *spatial feature* $\Upsilon_{\mathbf{x}}$, which is the median value computed on each pixel's adaptive neighborhood. The easiest way to use both pieces of information would be to build a stacked vector, but it would not allow the weighting of the different features. In our work, the kernel trick [72] of the SVM was exploited to design a composite kernel that allows the setting of the relative influence of the extracted features. This is detailed in Section IV-D.

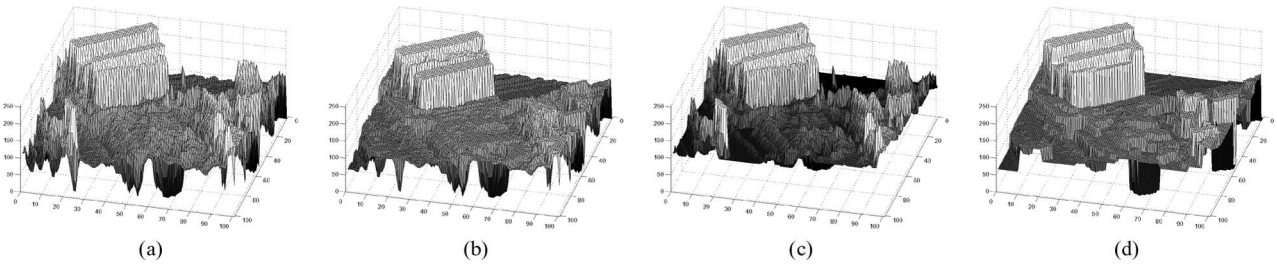

Fig. 7. *Limitations of the morphological profile. (a) Graph of the image in Fig. 5(a). (b) Graph of the geodesic opening of image Fig. 5(a). (c) Graph of the geodesic closing of image in Fig. 5(a). (d) Graph of the image in Fig. 5(a) filtered by the self-complementary area filter. From (b) and (c), it can be seen that only extrema are processed with the geodesic operators, while all the structures are processed on (d).*

D. Spectral–Spatial Classification

The SVM classifier has shown to be adapted to the classification of high-dimensional and/or multisource image [73], [74]. Furthermore, thanks to the kernel function, including many spatial features in the classification process is convenient. Several approaches were investigated for combining the spatial and spectral information in the classification process.

1) Feature Fusion: The EMP was originally used as an input to the classifier [40]. Good results in terms of classification accuracies were achieved. However, the EMP contains only a part of the spectral information from the data. To overcome this problem, data fusion was considered in [75]. The strategy uses both the EMP and the original hyperspectral image by combining them into a stacked vector. Furthermore, feature extraction could be also applied on both feature vectors and the extracted features are concatenated in one stacked vector and classified by an SVM classifier. It has been shown that SVM can suffer from the dimensionality if many features are irrelevant or redundant. However, the feature extraction can overcome the problem [76].

Noting $\mathbf{x}_\varphi$, the features associated to the spectral bands, and $\mathbf{x}_\omega$, the features associated to the EMP, the corresponding extracted features from the feature extraction algorithm are

$$\mathfrak{x}_\varphi = \Phi_\varphi^T \mathbf{x}_\varphi \tag{11}$$

and

$$\mathfrak{x}_\omega = \Phi_\omega^T \mathbf{x}_\omega \tag{12}$$

where Φ is the mapping matrix of the linear feature extraction algorithm. The stacked vector is constructed as $\mathfrak{x} = [\mathfrak{x}_\varphi, \mathfrak{x}_\omega]^T$. Note that, in this work, only morphological information was extracted, but it is possible to extract other types of spatial information with other processing and include them in the stacked vector.

2) Composite Kernel: Rather than building a stacked vector before the classification, it is possible to combine kernel functions to include both spatial and spectral classifications in the SVM classification process [67], [77], [78]. The linearity property was used to construct a spectral–spatial kernel $\mathcal{K}$, namely, the composite spectral–spatial kernel

$$\mathcal{K}_{\sigma,\mu}(\mathbf{x}, \mathbf{z}) = (1-\mu) k_\sigma^{\text{spat}}(\Upsilon_\mathbf{x}, \Upsilon_\mathbf{z}) + \mu k_\sigma^{\text{spect}}(\mathbf{x}, \mathbf{z}) \tag{13}$$

where σ is the width of the conventional Gaussian kernel

$$k_\sigma(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\sigma^2}\right) \tag{14}$$

and μ is a class-dependent weight parameter that controls the relative proportion of spatial and spectral information in the final kernel. For instance, for the class "grass," the spectral information should be more discriminative while spatial information should be more discriminative for the class "building." These hyperparameters are tuned during the training process of the SVM.

E. Experimental Evaluation of the Classification of the Morphological Features

In this section, the different classification strategies using the morphological approaches are compared. For each experiment, the EMP was built using the PCA and the KPCA. The number of (K)-principal components (PCs) selected explains 95% of the total variance. For both data sets, the three first PCs were selected. With the KPCA, for the University Area data set, the first 12 KPCs are needed to achieve 95% of the cumulative variance and 10 for the Pavia Center data set. A circular SE with a step size increment of 2 was used. Four openings and closings were computed for each (K)PC, resulting in an EMP of dimension $9 \times m$ [m being the number of retained (K)PCs]. For the feature fusion approach, several feature extraction techniques were investigated [75]. The DBFE and the NWFE provided good results in terms of classification accuracy (see Appendix B for a short description of the DBFE and the NWFE). For the computation of the morphological neighborhood, the area parameter was set to 30 for the University Area data set and to 20 for the Pavia Center data set. Note that there is a relatively large range of values for this parameter which provides good results in terms of accuracy; see [67]. Finally, all the hyperparameters of the SVM were selected using a fivefold cross validation [79].

The results are given in Tables 6 and 7. For the University Area data set, the best area parameter value for the area filtering was 30 and the best feature extraction method for the feature fusion approach was the DBFE with a threshold value on the cumulative variance of 95%. The classification results are significantly different, except the classification obtained with the spectral information only and the EMP ($Z < 1.96$). The best classification in terms of accuracy is obtained with the EMP built with the KPCA with a kappa equal to 0.95. The feature fusion with spectral–spatial feature extraction provide the second best results in terms of accuracy, with a kappa equal to 0.84. The third best kappa is 0.82 for the composite kernel approach.

Table 6 Classification Accuracies for University Area Data Set. The Best Results for Each Class Are Reported in Boldface. $\mathcal{K}^{\lambda}$ Means That Classification Was Performed Using the Composite Kernel and Area Filtering of Size λ, Spec-EMP Means That Classification Was Performed Using the Stacked Vector With the Spectral and the EMP, and DBFE-95% Means That Classification Was Performed Using the Extracted Spatial and Spectral Features Using DBFE and 95% of the Cumulative Variance

Class	spec	EMP-PCA	EMP-KPCA	$\mathcal{K}^{30}$	spec-EMP	DBFE-95%
1. Asphalt	83.7	94.2	**96.0**	83.6	95.1	90.5
2. Meadow	65.3	72.1	**97.5**	77.9	72.7	85.5
3. Gravel	68.4	45.9	81.1	**82.9**	60.6	51.3
4. Tree	97.9	98.8	**99.3**	96.7	99.1	99.2
5. Metal Sheet	99.4	**99.5**	99.4	98.7	99.4	99.4
6. Bare Soil	92.5	52.8	92.2	**95.2**	82.6	83.9
7. Bitumen	88.9	94.8	**98.8**	94.0	96.2	93.5
8. Brick	92.0	94.9	**99.4**	95.0	95.8	95.5
9. Shadow	96.0	89.6	**98.0**	97.4	92.5	94.2
OA	78.2	77.7	**96.3**	85.1	82.4	87.1
AA	87.1	82.5	**95.7**	91.2	88.2	88.1
κ	0.73	0.71	**0.95**	0.81	0.77	0.83

For the Pavia Center data set, the best area parameter value was 20. For this image, the NWFE was the best feature extraction method for the fusion approach. It provides, with the EMP–KPCA, the best results in terms of classification accuracy, but the difference between the two classifications is not significant ($Z < 1.96$). The second best result in terms of accuracy is given conjointly by the EMP–PCA and the feature fusion without feature extraction. Their classifications are very similar ($Z = 0.06$).

For both data sets, the use of the spatial information conjointly with the spectral information provides better classification results in terms of accuracy. For instance, for the University Area data set, the improvement of the global accuracy is about 20%. A small improvement (0.8%), but still significant, is observed for the Pavia Center data set, because the classification accuracy is already high using the spectral information only. However, the improvement corresponds to about 1185 additional correctly classified pixels. Also, the thematic maps are more homogeneous, as can be seen in Figs. 9 and 10. The "salt-and-paper" classification noise of the thematic map obtained with the spectral information alone is removed or reduced when adding the spatial information in the classification process. Last, it has been observed that when the number of training samples is limited, better classification results are obtained when combining the spatial and spectral information than using the spectral information only [75], [78].

F. Future Trends in Morphological Processing for the Spectral–Spatial Classification of Hyperspectral Images

Recently, new connected morphological operators have been investigated for the analysis of hyperspectral images. They are based on a tree-based image representation [80]. Attribute filters offer new possibilities for extracting morphological information [81]. They are able to filter the spatial structures according to their geometry (area, length, shape factors), texture (range, entropy), etc. [82]. It is possible to construct the EMP using the same methodology as with the conventional geodesic operators,

Table 7 Classification Accuracies for Pavia Center Data Set. The Best Results for Each Class Are Reported in Boldface. $\mathcal{K}^{\lambda}$ Means That Classification Was Performed Using the Composite Kernel and Area Filtering of Size λ, Spec-EMP Means That Classification Was Performed Using the Stacked Vector With the Spectral and the EMP, and DBFE-95% Means That Classification Was Performed Using the Extracted Spatial and Spectral Features Using NWFE and 95% of the Cumulative Variance

Class	spec	EMP-PCA	EMP-KPCA	$\mathcal{K}^{20}$	spec-EMP	NWFE-99%
1. Water	99.1	99.1	98.9	**99.2**	98.7	**99.2**
2. Tree	90.8	91.6	92.0	90.0	**93.5**	92.5
3. Meadow	97.4	96.2	96.3	**98.1**	96.0	96.8
4. Brick	87.5	98.4	**99.6**	94.0	98.8	**99.6**
5. Bare Soil	94.6	98.8	**99.8**	99.5	99.4	99.7
6. Asphalt	96.4	98.0	**99.2**	95.8	98.4	98.7
7. Bitumen	96.5	97.9	**98.6**	98.2	98.2	98.4
8. Tile	99.5	99.7	**99.9**	99.5	99.8	99.7
9. Shadow	**100**	99.4	99.6	99.9	99.9	96.9
OA	98.1	98.7	98.9	98.4	**99.7**	98.8
AA	95.1	97.7	**98.3**	97.1	98.1	97.5
κ	0.97	**0.98**	**0.98**	0.97	**0.98**	**0.98**

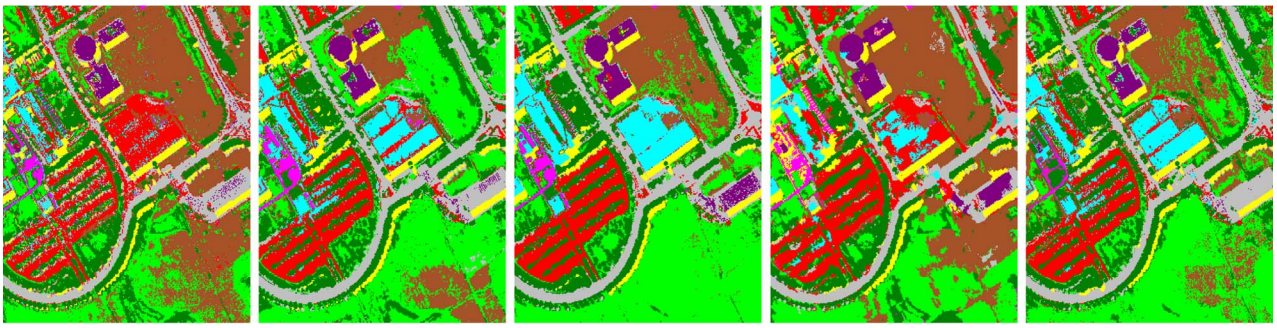

Fig. 9. ***Thematic maps obtained with the University Area data set: (a) spec, (b) EMP-PCA, (c) EMP-KPCA, (d) $\mathcal{K}^{30}$, and (e) DBFE-95%.***

as described in [83]. However, the definition of adapted attributes for a specific application is still an ongoing research.

The need for an ordering relation is still an important issue in morphological hyperspectral image processing. Valero *et al.* have proposed an alternative strategy based on a binary partition tree that allows the processing of the hyperspectral image without any feature reduction method [84]. The proposed representation is used for image simplification and segmentation. Surely, new possibilities in terms of morphological neighborhood can be offered and should be investigated in relation with the problem of classification. Similarly, the extension of self-complementary area filters to multivalued pixels is opening new paths for the characterization of the morphological neighborhood [85].

The spectral–spatial classification method could also benefit from recent work on multisource classification. For instance, the recently proposed multiple kernel learning (MKL) method may provide a nice framework to fuse the output of several attribute filters for the purpose of classification [86], [87]. However, the actual computational load of MKL algorithms makes them not well adapted for the classification of hyperspectral images.

V. SPATIAL REGULARIZATION OF PIXEL-WISE CLASSIFICATION USING SEGMENTATION

Even though the use of morphological profiles or area filters for spectral–spatial classification improves classification accuracies when compared to pixel-wise classification, these methods raise the problem of neighborhoods' scale selection. In this section, a spatial–spatial classification approach is presented using adaptive spatial neighborhoods derived from a segmentation map. First, three segmentation methods for hyperspectral images are discussed, and then an algorithm for combining the extracted spatial regions with spectral information into a classifier is presented.

Segmentation techniques can be grouped into three classes [88].

- Working in the spatial domain: These methods search for groups of spatially connected pixels, i.e., regions, which are similar according to the defined criterion. Examples are region growing, split-and-merge, and watershed techniques [43].
- Working in the spectral domain: These approaches search for similarities between image pixels and clusters of pixels, not taking into consideration the

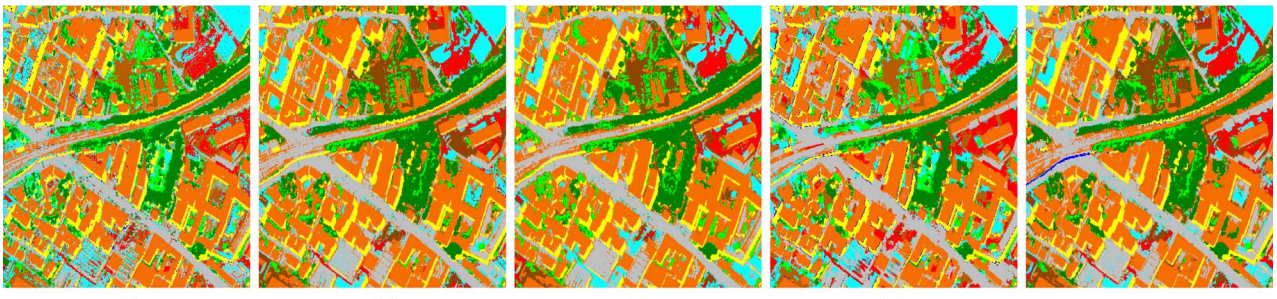

Fig. 10. ***Thematic maps obtained with the Center Pavia data set: (a) spec, (b) EMP-PCA, (c) EMP-KPCA, (d) $\mathcal{K}^{20}$, and (e) NWFE-99%.***

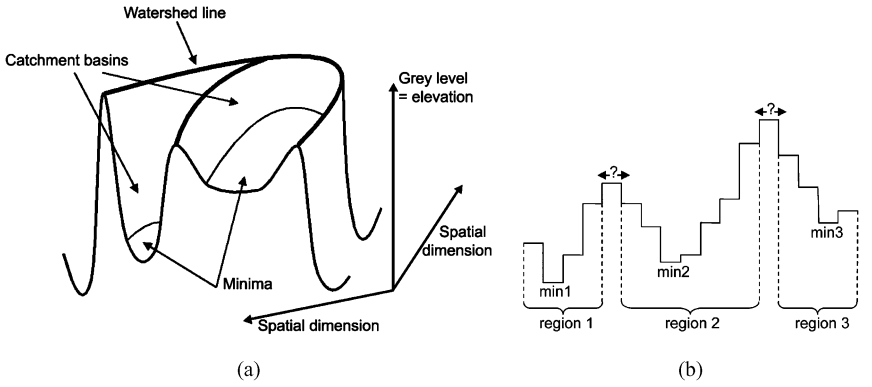

Fig. 11. ***(a) Topographic representation of a one-band image. (b) Example of a watershed transformation in 1-D. Illustration taken from [44].***

spatial location of these pixels. Segmentation map is obtained by a follow-up processing which allocates different labels for disjoint regions within the same cluster. Examples are thresholding and partitional clustering methods [88].

- Combining spatial-based and spectral-based segmentation. An example is an HSeg algorithm [89].

In the following, one technique from each class of segmentation methods is investigated: 1) spatial-based segmentation using watershed transformation; 2) spectral-based segmentation using expectation–maximization (EM) algorithm [90], [91]; and 3) segmentation in both spatial and spectral domains using the HSeg algorithm [89].

A. Watershed Segmentation

Watershed transformation is a powerful morphological approach for image segmentation which combines region growing and edge detection. It considers a 2-D one-band image as a topographic relief [48], [92]. The value h of a pixel stands for its elevation. The watershed lines divide the image into catchment basins, so that each basin is associated with one *minimum* in the image (see Fig. 11). The watershed is usually applied to the gradient function, and it divides an image into regions, so that each region is associated with one *minimum* of the gradient image.

As with morphological profile (see Section IV-B), the extension of a watershed technique to the case of hyperspectral images is not straightforward, because there is no natural means for total ordering of multivariate pixels. Several techniques for applying watershed to hyperspectral images have been proposed in [44] and [93]. The most common approach consists in computing a one-band gradient from a multiband image, and then executing a standard watershed algorithm. One such algorithm is presented in the following [44].

1) First, a one-band robust color morphological gradient (RCMG) [94] of a hyperspectral image is computed. For each d-band pixel vector $\mathbf{x}_p \in \mathbb{R}^d$, let $\chi = [\mathbf{x}_p^1, \mathbf{x}_p^2, \ldots, \mathbf{x}_p^e]$ be a set of e vectors contained within an SE B (i.e., the pixel $\mathbf{x}_p$ itself and $e-1$ neighboring pixels). A 3×3 square SE with the origin in its center is typically used. The color morphological gradient (CMG), using the Euclidean distance, is computed as

$$\mathrm{CMG}_B(\mathbf{x}_p) = \max_{i,j \in \chi} \left\{ \left\| \mathbf{x}_p^i - \mathbf{x}_p^j \right\|_2 \right\} \tag{15}$$

i.e., the *maximum* of the distances between all pairs of vectors in the set χ. One of the drawbacks of the CMG is that it is very sensitive to noise. In order to overcome the problem of outliers, the RCMG has been proposed [94]. The algorithm for making a CMG robust consists in removing the two pixels that are farthest apart and then finding the CMG of the remaining pixels. This process can be repeated several times depending on the size of an SE and noise level. Thus, the RCMG, using the Euclidean distance, can be defined as

$$\mathrm{RCMG}_B(\mathbf{x}_p) = \max_{i,j \in [\chi - REM_r]} \left\{ \left\| \mathbf{x}_p^i - \mathbf{x}_p^j \right\|_2 \right\} \tag{16}$$

where REM_r is a set of r vector pairs removed. If a 3×3 square SE is used, $r = 1$ is recommended [94].

2) Subsequently, the watershed transformation is applied on the one-band RCMG image, using a standard algorithm, for example, the algorithm of Vincent and Soille [95]. As a result, the image is segmented into a set of regions, and one subset of watershed pixels, i.e., pixels situated on the borders between regions (see Fig. 11).

3) Finally, every watershed pixel is assigned to the neighboring region with the "closest" median [71], i.e., with the minimal distance between the vector median of the corresponding region and the watershed pixel. Assuming that an L_1-norm is used to compute distances, a vector median for the region $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^d, j = 1, 2, \ldots, l\}$ is defined as $\mathbf{x}_{VM} = \arg\min_{\mathbf{x}\in\mathbf{X}}\{\sum_{j=1}^{l} \|\mathbf{x} - \mathbf{x}_j\|_1\}$.

B. Segmentation by EM

The EM algorithm for the Gaussian mixture resolving belongs to the class of techniques working in the spectral domain. It is a partitional clustering approach, which groups all the pixels into clusters of spectrally similar pixels [45], [90] [91]. The use of partitional clustering for hyperspectral image segmentation has been discussed in [45].

In the EM algorithm, it is assumed that pixels belonging to the same cluster are drawn from a multivariate Gaussian probability distribution. Each image pixel can be statistically modeled by the following probability density function:

$$p(\mathbf{x}) = \sum_{c=1}^{C} \omega_c \phi_c(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{17}$$

where C is the number of clusters, $\omega_c \in [0, 1]$ is the mixing proportion (weight) of a cluster c with $\sum_{c=1}^{C} \omega_c = 1$, and $\phi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$$\phi_c(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_c|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} (\mathbf{x} - \boldsymbol{\mu}_c)\right\}. \tag{18}$$

The distribution parameters $\psi = \{C, \omega_c, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c; c = 1, 2, \ldots, C\}$ are estimated using the iterative classification EM (CEM) algorithm, as described in [45] (see Appendix C). An upper bound on the number of clusters, which is a required input parameter, is recommended to be chosen slightly superior to the number of classes.

When the algorithm converges, the partitioning of the set of image pixels into C clusters is obtained. Because no spatial information is used during the clustering procedure, pixels with the same cluster label can either form a connected spatial region, or can belong to disjoint regions. In order to obtain a segmentation map, a connected components labeling algorithm [96] is applied to the cluster partitioning. This algorithm allocates different labels for disjoint regions within the same cluster.

The total number of parameters to be estimated by the EM algorithm is $P = (d(d+1)/2 + d + 1)C + 1$, where d is a dimensionality of feature vectors. If the value of d is large, P may be quite a large number. This may cause the problem of the covariance matrix singularity or inaccurate parameter estimation results. In order to avoid these problems, a feature reduction should be previously applied. The use of a piecewise constant function approximations method (PCFA) [97] has been investigated, which is a simple dimensionality reduction approach that has shown good performances for hyperspectral data feature extraction in terms of classification accuracies.

C. HSeg Segmentation

The HSeg algorithm is a segmentation technique combining region growing, using the hierarchical stepwise optimization (HSWO) method [98], which produces spatially connected regions, with unsupervised classification, that groups together similar spatially disjoint regions [89], [47]. The algorithm can be summarized as follows.

Initialization: Initialize the segmentation by assigning each pixel a region label. If a presegmentation is provided, label each pixel accordingly. Otherwise, label each pixel as a separate region.

1) Calculate the dissimilarity criterion value between all pairs of spatially adjacent regions. A spatially adjacent region for a given region is the one containing pixels situated in the neighborhood (e.g., eight-neighborhood) of the considered region's pixels.Different measures can be applied for computing dissimilarity criteria between regions, such as vector norms or spectral angle mapper (SAM) between the region mean vectors [47]. We present in this paper the use of the SAM criterion. The SAM measure between $\mathbf{x}_i$ and $\mathbf{x}_j$ $(\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d)$ determines the spectral similarity between two vectors by computing the angle between them. It is defined as

$$\mathrm{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\sum_{b=1}^{d} x_{ib} x_{jb}}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}\right). \tag{19}$$

2) Find the smallest dissimilarity criterion value dissim_val and set thresh_val equal to it. Then, merge all pairs of spatially adjacent regions with dissim_val = thresh_val.
3) If the parameter $S_{\mathrm{wght}} > 0.0$, merge all pairs of spatially nonadjacent regions with dissim_val $\leq S_{\mathrm{wght}} \cdot$ thresh_val.The optional parameter S_{wght} sets the relative importance of clustering based on spectral information only *versus* region growing. When $S_{\mathrm{wght}} = 0.0$, only spatially adjacent regions

are allowed to merge. When $0.0 < S_{\text{wght}} \leq 1.0$, spatially adjacent merges are favored compared with spatially nonadjacent merges by a factor of $1.0/S_{\text{wght}}$.

4) Stop if convergence is achieved. Otherwise, return to step 1.

Allowing for the merging of spatially disjoint regions leads to heavy computational demands. In order to reduce these demands, a recursive divide-and-conquer approximation of HSeg (RHSeg) and its efficient parallel implementation have been developed.

HSeg produces as output a hierarchical sequence of image segmentations from initialization down to the one-region segmentation, if allowed to proceed that far. In this sequence, a particular object can be represented by several regions at finer levels of details, and can be assimilated with other objects in one region at coarser levels of details. However, for practical applications, a subset of one or several segmentations needs to be selected out from this hierarchy. An appropriate level of segmentation detail can be chosen interactively with the program HSegViewer [47], or an automated method, tailored to the application, can be developed, such as explored in [100]–[102].

D. Spectral–Spatial Classification Using Majority Voting

Once image segmentation is performed, the next step is to incorporate the spatial information derived from a segmentation map in spectral–spatial classification. Different approaches of combining spatial and spectral information for classification have been proposed in the state of the art. Widayati *et al.* [103] and Linden *et al.* [104] applied an object-based classification approach, which consisted in assigning each region from the segmentation map to one of the classes using its vector mean as a feature. Experimental results proved that the representation of each region by its vector mean alone yields in most cases to spectral and textural information loss, resulting in imprecisions of classification. An alternative type of spectral–spatial classification consists in combining both spectral and spatial information within a feature vector of each pixel, and then classifying each pixel using these feature vectors. This method was described and investigated in Section IV, using either stacked features or composite kernels.

In this section, another classification approach is proposed, called *majority vote* [45].[5]

1) A pixel-wise classification, based on spectral information of pixels only, and a segmentation are independently performed. It is proposed to use an SVM pixel-wise classifier, which efficiently handles hyperspectral data.
2) For every region in the segmentation map, all the pixels are assigned to the most frequent class within this region.

[5]In the literature, this approach is often referred to as *plurality vote*.

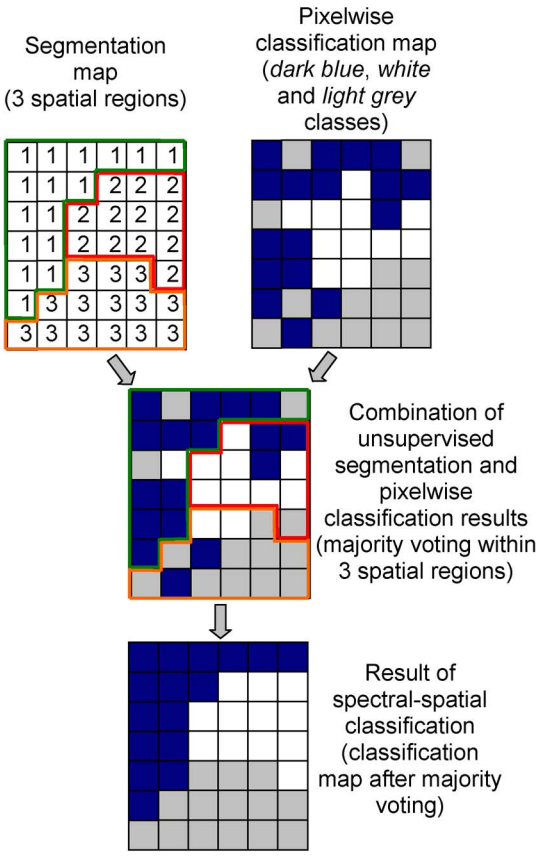

Fig. 12. ***Schematic example of spectral–spatial classification using majority voting within segmentation regions. Illustration taken from [46].***

Fig. 12 shows an illustrative example of the combination of spectral and spatial information using the majority voting classification method. The described approach retains all the spectral information for accurate image classification with a well-suited technique, while not increasing data dimensionality. Thus, it has proven to be an accurate, simple, and fast technique. Experimental results for the presented spectral–spatial classification approach using segmentation are presented in Section VI.

VI. SEGMENTATION AND CLASSIFICATION USING AUTOMATICALLY SELECTED MARKERS

As mentioned earlier, accurate segmentation results depend on the chosen measure of a region homogeneity, which is application specific [43]. If the final objective is to compute a supervised classification map, the information about thematic classes can be exploited for building a segmentation map. In this section, marker-controlled segmentation is explored, where markers for spatial regions are automatically derived from probabilistic

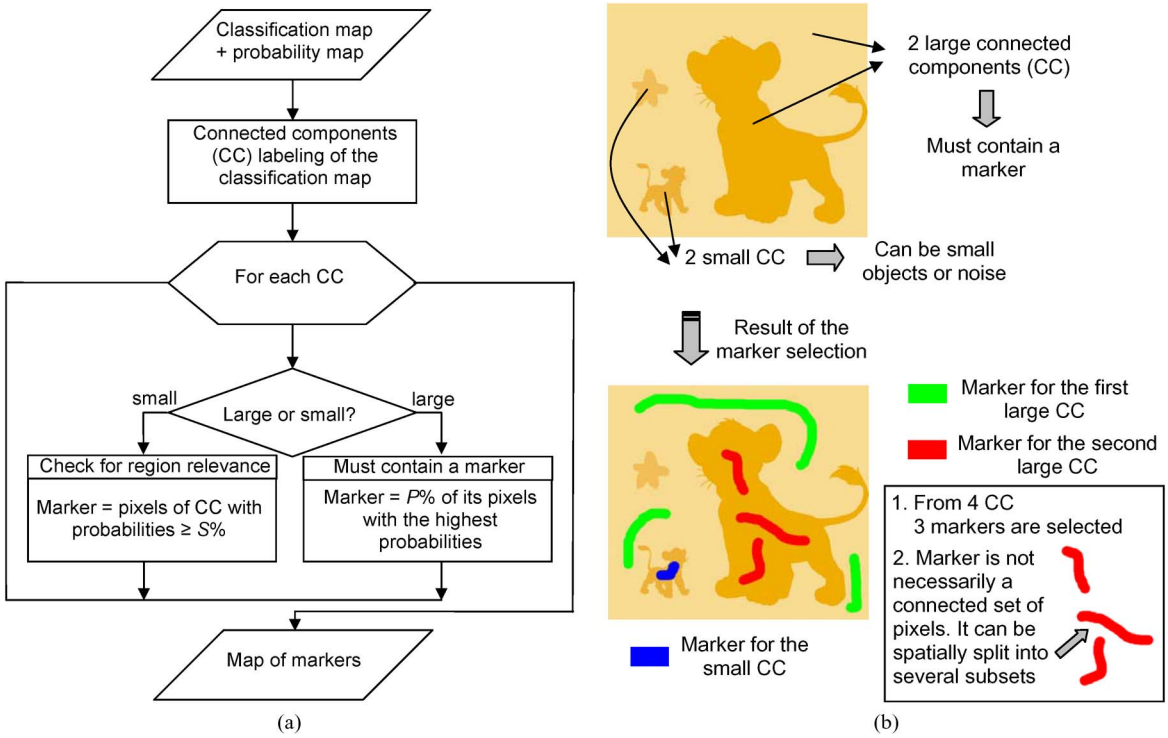

Fig. 13. ***(a) Flowchart of the SVM-based marker selection procedure. (b) Illustrative example of the SVM-based marker selection. Illustration taken from [49].***

classification results and then used as seeds for region growing [46], [49]. Assuming that classification results are typically more accurate inside spatial regions and more erroneous closer to region borders, it is proposed to choose the most reliably classified pixels as region markers. Two different marker selection approaches are presented further, based either on results of probabilistic SVM or an MC system. Then, a marker-controlled segmentation algorithm is described which consists in the construction of an MSF rooted on markers.

A. Marker Selection Using Probabilistic SVM

In [49], Tarabalka *et al.* choose markers by analyzing probabilistic SVM classification results. The proposed marker selection method consists of two steps (see the flowchart and the illustrative example in Fig. 13).

1) *Probabilistic pixel-wise classification*: Apply a probabilistic pixel-wise SVM classification of a hyperspectral image [72], [105]. The outputs of this step are a classification map, containing a unique class label for each pixel, and a probability map, containing probability estimates for each pixel to belong to the assigned class.In order to compute class probability estimates, pairwise coupling of binary probability estimates can be applied [105], [106]. In our work, the probabilistic SVM algorithm implemented in the LIBSVM library [105] was used. The objective is to estimate, for each pixel $\mathbf{x}$, classification probabilities

$$p(y|\mathbf{x}) = \{p_i = p(y = c|\mathbf{x}), i = 1, \ldots, K\} \quad (20)$$

where C is a number of thematic classes. For this purpose, pairwise class probabilities $r_{ij} \approx p(y = i|y = i \text{ or } j, \mathbf{x})$ are first estimated. Then, the probabilities in (20) are computed, as described in [106]. A probability map is further built by assigning to each pixel the *maximum* probability estimate $\max(p_i)$, $i = 1, \ldots, K$.

2) *Marker selection*: Perform a connected component labeling on the classification map, using an eight-neighborhood connectivity [96]. Then, analyze each connected component.
   - If a region is *large*, i.e., a number of pixels in the region $> M$, it is considered to represent a spatial structure. Its marker is defined as the $P\%$ of pixels within this region with the highest probability estimates.
   - If a region is *small*, it is further investigated if its pixels were classified to a particular class with a high probability. Otherwise, the component is assumed to be the consequence of

classification noise, and the algorithm tends to eliminate it. Its potential marker is formed by the pixels with probability estimates higher than a defined threshold S.

The procedure of the setting of parameters (M, P, S) based on *a priori* information for the image is described in [49]:

- A parameter M, which is a threshold of the number of pixels defining if the region is large, depends on the resolution of the image and typical sizes of the objects of interest.
- A parameter P, defining the percentage of pixels within the large region to be used as markers, depends on the previous parameter. Because the marker for a large region must have at least one pixel, the following condition must be fulfilled: $P \geq 100\%/M$.
- A parameter S, which is a threshold of probability estimates defining potential markers for small regions, depends on the probability of the presence of small structures in the image (which depends on the image resolution and the classes of interest), and the importance of the potential small structures (i.e., the cost of losing the small structures in the classification map).

At the output of the marker selection step, a map of m markers is obtained, where each marker $\mathbf{O}_i = \{\mathbf{x}_j \in \mathbf{X}, j = 1, \ldots, \mathrm{card}(\mathbf{O}_i); y_{\mathbf{O}_i}\}$ $(i = 1, \ldots, m)$ consists of one or several pixels and has a class label $y_{\mathbf{O}_i}$. One should note that a marker is not necessarily a spatially connected set of pixels.

B. Multiple-Classifier Approach for Marker Selection

Although the previously described marker selection approach has shown good results, the drawback of this method is that the choice of markers strongly depends on the performances of the selected pixel-wise classifier (e.g., the SVM classifier). In order to mitigate this dependence, it is proposed to use not a single classification algorithm for marker selection, but an ensemble of classifiers, i.e., multiple classifiers (MCs) [46]. For this purpose, several individual classifiers are combined within one system (see Fig. 14) in such a way that the complementary benefits of each classifier are exploited, while their weaknesses are avoided [107]. Fig. 15 shows a flowchart of the proposed multiple spectral–spatial classifier (MSSC) marker selection scheme, which consists of the following two steps.

1) *Multiple classification*: Apply several individual classifiers to an image. Spectral–spatial classifiers are used as individual classifiers for the MC system, each of them combining the results of a pixel-wise classification and one of the unsupervised segmentation techniques. The procedure is as follows:
   a) *Unsupervised image segmentation*: Segmentation methods based on different principles must be chosen. Three techniques described in Section V (watershed, segmentation by EM, and HSeg) are considered.
   b) *Pixel-wise classification*: The SVM method was used for classifying a hyperspectral image. This step results in a classification map, where each pixel has a unique class label.
   c) *Majority voting within segmentation regions*: Each of the obtained segmentation maps is combined with the pixel-wise classification map using the majority voting principle: for every region in the segmentation map, all the pixels are assigned to the most frequent class within this region (see Section V-D). Thus, q segmentation maps combined with the pixel-wise classification map result in q spectral–spatial classification maps.

Different segmentation methods based on dissimilar principles lead to different classification maps. It is important to obtain different results for an efficient MC system, so that potential mistakes of any given individual classifier get a chance to be corrected thanks to the complementary contributions of the other classifiers. By using spectral–spatial classifiers in this step, spatial context in the image is taken into account, yielding more accurate classification maps when compared with pixel-wise classification maps.

2) *Marker selection*: Another important issue for designing an MC system is the rule for combining the individual classifiers, i.e., the combination function [108]. The following exclusionary combination rule was proposed: for every pixel, if all the classifiers agree, keep this pixels as a marker, with the corresponding class label. The resulting map of m markers contains the most reliably classified pixels. The rest of the pixels are further classified by performing a marker-controlled region growing, as described in the following.

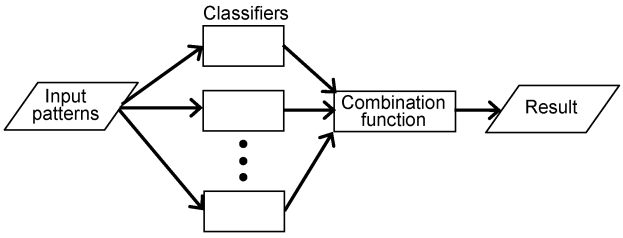

Fig. 14. *Flowchart of an MC system. Illustration taken from [46].*

C. Construction of an MSF

Once marker selection is performed, the obtained map of markers is further used for marker-controlled region growing, based on an MSF algorithm [46], [49]. The flowchart of the spectral–spatial classification using an

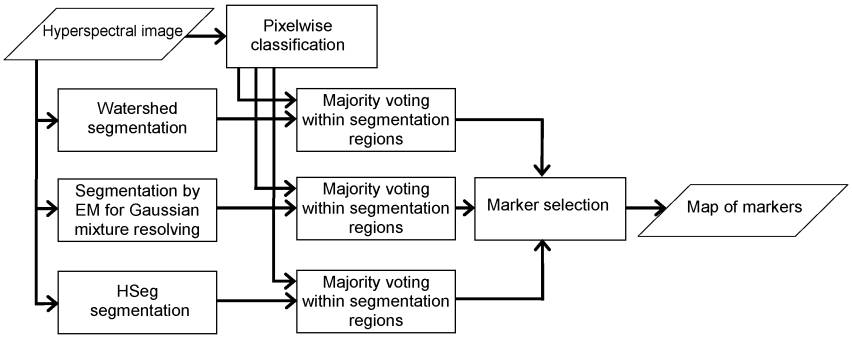

Fig. 15. ***Flowchart of the MSSC marker selection scheme.***

MSF grown from the classification-derived markers is depicted in Fig. 16. In the following, the two steps of the proposed procedure are described: construction of an MSF and majority voting within connected components.

1) *Construction of an MSF*: Each image pixel is considered as a vertex $v \in V$ of an undirected graph $G = (V, E, W)$, where V and E are the sets of vertices and edges, respectively, and W is a weighting function. Each edge $e_{i,j} \in E$ of this graph connects a couple of vertices i and j corresponding to the neighboring pixels. An eight-neighborhood was assumed in our work. A weight $w_{i,j}$ is assigned to each edge $e_{i,j}$, which indicates the degree of dissimilarity between two vertices connected by this edge. Different dissimilarity measures can be used for computing weights of edges, such as vector norms and SAM between two pixel vectors.

   Given a graph $G = (V, E, W)$, a spanning forest $F = (V, E_F)$ of G is a nonconnected graph without cycles such that $E_F \subset E$. The MSF rooted on a set of m distinct vertices $\{t_1, \ldots, t_m\}$ is defined as a spanning forest $F^* = (V, E_{F^*})$ of G, such that each tree of F^* is grown from one root t_i, and the sum of the edges weights of F^* is minimal [109]

$$F^* \in \arg\min_{F \in SF} \left\{ \sum_{e_{i,j} \in E_F} w_{i,j} \right\} \tag{21}$$

where SF is a set of all spanning forests of G rooted on $\{t_1, \ldots, t_m\}$.For constructing an MSF rooted on markers, m extra vertices t_i, $i = 1, \ldots, m$, are introduced. Each additional vertex t_i is connected by the null-weight edge with the pixels belonging to the marker $\mathbf{O}_i$. Furthermore, a root vertex r is added and is connected by the null-weight edges to the vertices t_i (Fig. 17 shows an example of addition of extra vertices). The *minimum* spanning tree [109] of the built graph induces an MSF in G, where each tree is grown on a vertex t_i. Prim's algorithm can be applied for computing a *minimum* spanning tree (See Appendix D) [49], [110]. The MSF is obtained after removing the vertex r. Each tree in the MSF forms a region in the segmentation map, by mapping the output graph onto an image. Finally, a spectral–spatial classification

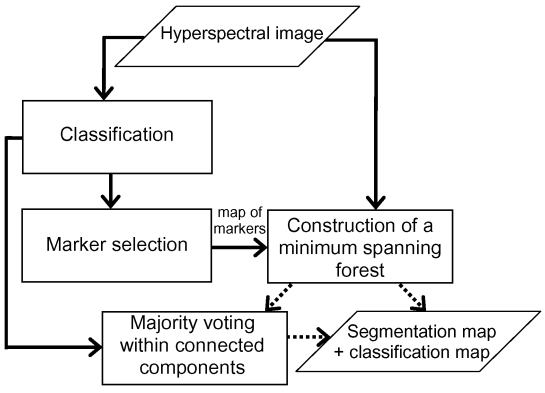

Fig. 16. ***Flowchart of the spectral–spatial classification approach using an MSF grown from automatically selected markers.***

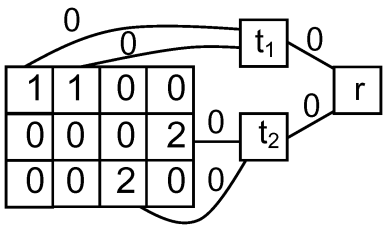

Fig. 17. ***Example of addition of extra vertices t_1, t_2, r to the image graph for construction of an MSF rooted on markers 1 and 2. Nonmarker pixels are denoted by "0."***

Table 8 Classification Accuracies in Percentage for the University Area Data Set: Overall Accuracy (OA), Average Accuracy (AA), Kappa Coefficient (κ), and Class-Specific Accuracies

Class	WH+MV	EM+MV	HSeg+MV	SVMMSF	SVMMSF+MV	MSSC-MSF
1. Asphalt	93.6	90.1	94.8	93.1	93.2	**98.0**
2. Meadow	75.1	96.0	89.3	72.3	85.7	**96.7**
3. Gravel	66.1	82.3	96.1	89.2	89.2	**97.8**
4. Tree	98.6	85.5	98.1	87.0	91.2	**98.8**
5. Metal sheet	99.9	**100**	99.8	99.9	99.91	99.9
6. Bare soil	97.4	96.7	99.8	97.1	99.9	**100**
7. Bitumen	96.3	91.9	**100**	98.6	98.57	99.9
8. Brick	97.9	98.4	99.3	95.7	99.0	**99.8**
9. Shadow	97.0	97.4	96.5	**98.4**	96.2	96.5
OA	85.4	94.0	93.9	84.14	91.1	**97.9**
AA	91.3	93.1	97.0	92.35	94.8	**98.6**
κ	0.81	0.92	0.92	0.80	0.88	**0.97**

map is obtained by assigning the class of each marker to all the pixels grown from this marker.

2) *Majority voting within connected components (optional step)*: Although the most reliably classified pixels are selected as markers, it may happen that a marker is assigned to the wrong class. In this case, all the pixels within the region grown from this marker risk being wrongly classified. In order to make the proposed classification scheme more robust, the classification map can be postprocessed by applying a simple *majority voting* technique [45], [103]. For this purpose, connected component labeling is applied on the obtained spectral–spatial classification map, using a four-neighborhood connectivity. Then, for every connected component, all the pixels are assigned to the majority class when analyzing a pixel-wise classification map within this region.Note that an eight-neighborhood connectivity was used for building an MSF and a four-neighborhood connectivity for majority voting. The use of the eight-neighborhood connectivity in the first case enables one to obtain a segmentation map without rough borders. When performing the majority voting step, the use of the four-neighborhood connectivity results in the larger or the same number of connected components as the use of the eight-neighborhood connectivity. Hence, possible undersegmentation can be corrected in this step. One region from a segmentation map can be split into two connected regions when using the four-neighborhood connectivity. Furthermore, these two regions can be assigned to two different classes by the majority voting procedure.

D. Experimental Evaluation of Spectral–Spatial Classification Methods Using Segmentation-Derived Neighborhoods

In this section, spectral–spatial classification strategies described in Sections V and VI are compared. Tables 8 and 9 summarize both class-specific and global

Table 9 Classification Accuracies in Percentage for the Indian Pines Image Data Set: Overall Accuracy (OA), Average Accuracy (AA), Kappa Coefficient (κ), and Class-Specific Accuracies

Class	SVM	WH+MV	EM+MV	HSeg+MV	SVM-MSF	SVM-MSF+MV	MSSC-MSF
1. Corn-no till	78.2	**94.2**	89.1	90.5	91.0	93.2	89.7
2. Corn-min till	69.6	78.1	75.6	83.0	69.5	**96.6**	87.0
3. Corn	91.9	88.6	65.2	**95.7**	**95.7**	**95.7**	95.1
4. Soybeans-no till	82.0	**96.3**	88.1	92.1	98.0	93.9	91.8
5. Soybeans-min till	58.0	68.8	65.7	84.1	82.0	82.0	**89.2**
6. Soybeans-clean till	87.9	90.8	95.0	95.4	86.0	97.2	**97.4**
7. Alfalfa	74.4	**94.9**	**94.9**	92.3	**94.9**	**94.9**	**94.9**
8. Grass/pasture	92.2	**95.1**	94.0	94.4	94.6	94.6	94.6
9. Grass/trees	91.7	**98.0**	96.4	97.6	92.4	97.3	97.9
10. Grass/pasture-mowed	**100**	**100**	**100**	**100**	**100**	**100**	**100**
11. Hay-windrowed	97.7	99.5	99.3	99.5	**99.8**	**99.8**	**99.8**
12. Oats	**100**	**100**	40.0	**100**	**100**	**100**	**100**
13. Wheat	98.8	**99.4**	98.8	98.2	**99.4**	**99.4**	**99.4**
14. Woods	93.0	97.2	96.7	98.6	97.6	**99.7**	99.4
15. Bldg-Grass-Tree-Drives	61.5	69.4	66.7	**82.1**	68.8	68.8	73.6
16. Stone-steel towers	97.8	95.6	**100**	**100**	95.6	95.6	97.8
OA	78.2	86.6	83.6	90.8	88.4	91.8	**92.3**
AA	86.0	91.6	85.3	94.0	91.6	**94.3**	94.2
κ	0.75	0.85	0.81	0.90	0.87	0.91	**0.91**

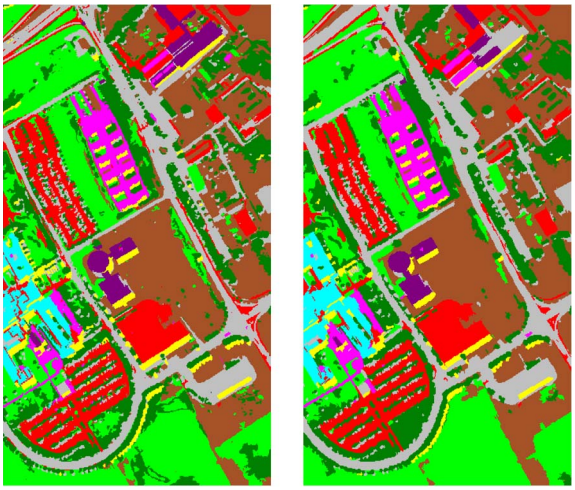

Fig. 18. ***Classification maps for the University Area data set. (a) HSeg+MV. (b) MSSC-MSF.***

accuracies of classification of the University Area and the Indian Pines data sets, respectively, using: 1) segmentation followed by majority voting (WH+MV, EM+MV, and HSeg+MV methods, using watershed, EM and HSeg segmentation, respectively); 2) marker selection using probabilistic SVM followed by MSF segmentation, without (SVMMSF method) and with (SVMMSF+MV method) optional majority voting step; and 3) marker selection using MSSC approach followed by MSF segmentation without the optional majority voting step (MSSC–MSF technique). Some of the corresponding classification maps are given in Figs. 18 and 19. Parameters for these methods were chosen following advice in [46] and [49].

- For the EM segmentation, a feature extraction was applied using the PCFA method to get a ten-band image. The *maximum* number of clusters was chosen to be equal to 10 and 17 for the University Area and Indian Pines images, respectively (typically slightly superior to the number of classes).
- For the HSeg algorithm, the parameters were tuned as $S_{\mathrm{wght}} = 0.1$ and $S_{\mathrm{wght}} = 0.0$ for the University Area and Indian Pines data sets, respectively. The reason for that is that while the former image contains spectrally dissimilar classes, the latter agricultural image has classes with very similar spectral responses, and best merge growing of adjacent regions yields the most accurate segmentation results for the latter image.
- For marker selection using probabilistic SVM, $M = 20$ and $P = 5$. In order to define a threshold S, the probability estimates for the whole image were sorted, and S was chosen equal to the lowest probability within the highest 2% of probability estimates.
- As recommended in [49], for the SVMMSF and SVMMSF+MV methods, the SAM dissimilarity measure was used for the Indian Pines image, and L_1 vector norm dissimilarity measure for the University Area image (for urban images containing shadows vector norms give better accuracies when compared with the SAM measure), respectively.
- As proposed in [46], the SAM dissimilarity measure is used for construction of an MSF in the MSSC-MSF technique.

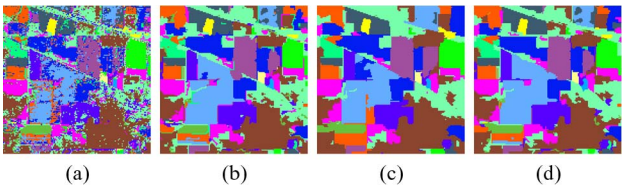

Fig. 19. ***Classification maps for the Indian Pines data set. (a) SVM. (b) HSeg+MV. (c) SVM-MSF+MV. (d) MSSC-MSF.***

As can be seen from Tables 8 and 9 and Figs. 18 and 19 (and compared to the results in Table 6), all the global spectral–spatial classification accuracies are higher when compared with the pixel-wise accuracies. The MSSC-MSF method yields the best overall accuracies. The Z test computed between the MSSC–MSF and the EMP–KPCA is positive for the MSSC–MSF ($Z = 2.82$). Thus, it is advantageous to apply segmentation techniques for extracting spatial dependencies in remote sensing images for the final objective of thematic classification. The segmentation has proven to be more accurate when incorporating additional class-specific information in a segmentation procedure, by means of introducing classification-derived markers for marker-controlled region growing. Spectral–spatial classification also benefits from the use of MC approaches, both for classification [107], [111] and marker selection [46].

VII. CONCLUSION

In this paper, spectral–spatial classification of hyperspectral images is addressed. Taking into account the need of spatial information during the classification process and the number of spectral components, several approaches were considered. The framework of the proposed methods can be summed up as extraction of spatial and spectral information and the combination of information either during the classification step or after a primary classification.

The extraction of the spatial features is done at the object level, providing more informative and more adaptive features. Morphological processing was used to perform a multiscale analysis of the interpixel dependency and to compute the morphological neighborhood for each pixel of the image. Another considered approach to compute adaptive neighborhoods consists in using regions

derived from a segmentation map. Several segmentation techniques for hyperspectral images were investigated. Hierarchical segmentation provided the most accurate segmentation map.

For the classification step, the SVM was used because of its capability to deal with high-dimensional data. Its flexibility, due to the kernel function, allows several strategies for including spatial features in the classification process: feature fusion or composite kernels. The first technique provides the best results in terms of classification accuracy, but the second one should be investigated deeper with new machine learning tools such as multiple kernel learning. An MC system was considered for combining segmentation and classification procedures.

For the three considered data sets, the classification accuracy is improved by the proposed methods and the resulting thematic maps are more homogeneous and spatially consistent. Two algorithms, the EMP–KPCA and the MSSC–MSF, provide leading performances in terms of classification accuracies.

Final advice for building an accurate classification system for hyperspectral images is as follows.

- Build an MC system.
- Classifiers should be robust to the dimensionality (e.g., SVM) with different inputs: spectral, spatial, and spectral–spatial.
- Use spatial regularization driven by presegmentation map. ■

APPENDIX I
ASSESSING THE ACCURACY

The classification accuracy was assessed with the overall accuracy (OA) which is the number of accurately classified samples divided by the number of test samples, the average accuracy (AA) which represents the average of class classification accuracy, the kappa coefficient of agreement (κ) which is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone, and the class-specific accuracy. These criteria were used to compare classification results and were computed using the confusion matrix.

Furthermore, the statistical significance of differences was computed using McNemar's test, which is based upon the standardized normal test statistic [112]

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}$$

where f_{12} indicates the number of samples classified correctly by classifier 1 and wrongly by classifier 2. At the commonly used 5% level of significance, the difference in accuracy between classifiers 1 and 2 is said to be statistically significant if $|Z| > 1.96$. The sign of Z indicates whether classifier 1 is more accurate than classifier 2 ($Z > 0$) or *vice versa* ($Z < 0$).

APPENDIX II
DBFE AND NWFE

The DBFE was proposed by Lee and Landgrebe [113] for the purpose of classification. From the decision boundary (the line in the feature space where a sample belongs equally to the classes), the DBFE permits the extraction of informative feature and permits the deletion of the redundant features. The decision boundary is found with a Gaussian mixture model, i.e., each class is modeled by a Gaussian distribution. The DBFE requires the estimation of the parameter (mean vector and covariance matrix). Hence, with limited training set, the method does not perform well.

To overcome these limitations, Kuo and Landgrebe have proposed the NWFE [114]. It is a nonparametric version of the linear discriminant analysis. The between- and within-class matrices are estimated in a nonparametric way, putting different weights on every sample to compute the local means. Hence, contrary to the DBFE, the NWFE seeks features that maximize the separability of the classes.

APPENDIX III
CEM CLUSTERING

Inputs:

- a set of n feature vectors (patterns) $\mathbf{X}$;
- an upper bound $C_{\max}$ on the number of clusters.

Initialization (Iteration 0):
Let $C = C_{\max}$. Determine the first partition $\mathbf{Q}_c^0$, $c = 1, 2, \ldots, C$, of $\mathbf{X}$:

1) Choose randomly C patterns from the set $\mathbf{X}$ to serve as cluster centers.
2) Assign the remaining patterns to the clusters on the basis of the nearest Euclidean distance to the cluster center.

For every iteration $i > 0$ (I iterations in total):
Parameter estimation step:
Estimate $\boldsymbol{\mu}_c^i$, $\boldsymbol{\Sigma}_c^i$, and ω_c^i for $c = 1, 2, \ldots, C$ by component-wise empirical means, empirical covariances, and relative frequencies, respectively [90]

$$\boldsymbol{\mu}_c^i = \frac{1}{m_c^{i-1}} \sum_{j=1}^{m_c^{i-1}} \mathbf{x}_{j,c}^{i-1} \tag{22}$$

$$\boldsymbol{\Sigma}_c^i = \frac{1}{m_c^{i-1}} \sum_{j=1}^{m_c^{i-1}} \left(\mathbf{x}_{j,c}^{i-1} - \boldsymbol{\mu}_c^i\right)\left(\mathbf{x}_{j,c}^{i-1} - \boldsymbol{\mu}_c^i\right)^T \tag{23}$$

$$\omega_c^i = \frac{m_c^{i-1}}{n}. \tag{24}$$

Cluster assignment step:

1) Assign each pattern in $\mathbf{X}$ to one of the clusters according to the maximum *a posteriori* probability criteria

$$\mathbf{x}_j \in \mathbf{Q}_c^i : p(c|\mathbf{x}_j) = \max_l p(l|\mathbf{x}_j) \tag{25}$$

where

$$p(c|\mathbf{x}_j) = \frac{\omega_c^i \phi_c\left(\mathbf{x}_j; \boldsymbol{\mu}_c^i, \boldsymbol{\Sigma}_c^i\right)}{\sum_{c=1}^{C} \omega_c^i \phi_c\left(\mathbf{x}_j; \boldsymbol{\mu}_c^i, \boldsymbol{\Sigma}_c^i\right)}. \tag{26}$$

2) Eliminate cluster c if m_c^i is less than the dimensionality of patterns, $c = 1, 2, \ldots, C$. The patterns that belonged to the deleted clusters will be reassigned to the other clusters in the next iteration.

3) If the convergence criterion is not achieved, return to the parameter estimation step.

APPENDIX IV
PRIM'S ALGORITHM

Require: Connected graph $G = (V, E, W)$
Ensure: Tree $T^* = (V^*, E^*, W^*)$
$V^* = \{v\}$, v is an arbitrary vertex from V
while $V^* \neq V$ **do**
Choose edge $e_{i,j} \in E$ with minimal weight such that $i \in V^*$ and $j \notin V^*$
$V^* = V^* \cup \{j\}$
$E^* = E^* \cup \{e_{i,j}\}$
end while

REFERENCES

[1] M. A. Cochrane, "Using vegetation reflectance variability for species level classification of hyperspectral data," *Int. J. Remote Sens.*, vol. 21, no. 10, pp. 2075–2087, 2000.

[2] A. Ghiyamat and H. Shafri, "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment," *Int. J. Remote Sens.*, vol. 31, no. 7, pp. 1837–1856, 2010.

[3] J. Pontius, M. Martin, L. Plourde, and R. Hallett, "Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2665–2676, 2008.

[4] E. A. Cloutis, "Hyperspectral geological remote sensing: Evaluation of analytical techniques," *Int. J. Remote Sens.*, vol. 17, no. 12, pp. 2215–2242, 1996.

[5] T. Schmid, M. Koch, and J. Gumuzzio, "Multisensor approach to determine changes of wetland characteristics in semiarid environments (central Spain)," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2516–2525, Nov. 2005.

[6] Y. Lanthier, A. Bannari, D. Haboudane, J. R. Miller, and N. Tremblay, "Hyperspectral data segmentation and classification in precision agriculture: A multi-scale analysis," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2008, pp. 585–588.

[7] J. L. Boggs, T. D. Tsegaye, T. L. Coleman, K. C. Reddy, and A. Fahsi, "Relationship between hyperspectral reflectance, soil nitrate-nitrogen, cotton leaf chlorophyll, and cotton yield: A step toward precision agriculture," *J. Sustainable Agriculture*, vol. 22, no. 3, pp. 5–16, 2003.

[8] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[9] X. Briottet, Y. Boucher, A. Dimmeler, A. Malaplate, A. Cini, M. Diani, H. Bekman, P. Schwering, T. Skauli, I. Kasen, I. Renhorn, L. Klasén, M. Gilmore, and D. Oxford, "Military applications of hyperspectral imagery," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6239, Jun. 2006, 62390B.

[10] D. A. Landgrebe, "Multispectral land sensing: Where from, where to?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 414–421, Mar. 2005.

[11] C. Chang, *Hyperspectral Imaging. Techniques for Spectral Detection and Classification*. Norwell, MA: Kluwer, 2003.

[12] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.

[13] M. G. Kendall, *A Course in the Geometry of n-Dimensions*. New York: Dover, 1961.

[14] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 28, no. 1, pp. 39–54, Feb. 1998.

[15] D. L. Donoho, "High-dimensional data analysis: The curses and blessing of dimensionality," in *AMS Math. Challenges 21st Century*, pp. 1–32, 2000.

[16] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[17] C. J. C. Burges, "Dimension reduction: A guided tour," *Found. Trends Mach. Learn.*, vol. 2, no. 4, pp. 275–365, 2010.

[18] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[19] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.

[20] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*. New York: Wiley, 2009.

[21] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2006, vol. 2, p. II, DOI: 10.1109/ICASSP.2006.1660467.

[22] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[23] S. Tadjudin and D. A. Landgrebe, "Classification of high dimensional data with limited training samples," Schl. Electr. Comput. Eng., Purdue Univ., Tech. Rep., 1998.

[24] J. Chanussot, J. A. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geoscience Remote Sens. Lett.*, vol. 3, no. 1, pp. 40–44, Jan. 2006.

[25] G. Martin and A. Plaza, "Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 52, pp. 380–395, 2012.

[26] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. 14, no. 1, pp. 19–26, Jan. 1976.

[27] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[28] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[29] X. Descombes, M. Sigelle, and F. Preteux, "GMRF parameter estimation in a non-stationary framework by a renormalization technique: Application to remote sensing imaging," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 490–503, Apr. 1999.

[30] X. Jia and J. A. Richards, "Managing the spectral-spatial mix in context classification using Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 311–314, Apr. 2008.

[31] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient ND image segmentation," *Int. J. Comput. Vis.*, vol. 70, pp. 109–131, Nov. 2006.

[32] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[33] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "SAR image regularization with fast approximate discrete minimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1588–1600, Jul. 2009.

[34] S. Le Hegarat-Mascle, A. Kallel, and X. Descombes, "Ant colony optimization for image regularization based on a nonstationary Markov modeling," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 865–878, Mar. 2007.

[35] Z. Bing, L. Shanshan, J. Xiuping, G. Lianru, and P. Man, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.

[36] R. C. González and R. E. Woods, *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 2008.

[37] S. Aksoy, "Spatial techniques for image classification," in *Signal and Image Processing for Remote Sensing*. London, U.K.: Taylor & Francis, 2006, pp. 491–513.

[38] G. Zhang, X. Jia, and N. M. Kwok, "Spectral-spatial based super pixel remote sensing image classification," in *Proc. 4th Int. Congr. Image Signal Process.*, Oct. 2011, vol. 3, pp. 1680–1684.

[39] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[40] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[41] J. A. Richards and X. Jia, "A Dempster-Shafer relaxation approach to context classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1422–1431, May 2007.

[42] B. Zhang, X. Jia, Z. Chen, and Q. Tong, "A patch-based image classification by integrating hyperspectral data with GIS," *Int. J. Remote Sens.*, vol. 27, no. 15, pp. 3337–3346, 2006.

[43] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[44] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[45] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 2973–2987, Sep. 2009.

[46] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[47] J. C. Tilton, Y. Tarabalka, P. M. Montesano, and E. Gofman, "Best merge region growing segmentation with integrated non-adjacent region object aggregation," *IEEE Trans. Geosci. Remote Sens.*, 2012, DOI: 10.1109/TGRS.2012.2190079.

[48] P. Soille, *Morphological Image Analysis, Principles and Applications*, 2nd ed. New York: Springer-Verlag, 2003.

[49] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst. Man Cybern. B, Cybern*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[50] S. Tadjudin and D. A. Landgrebe, "Covariance estimation with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 4, pp. 2113–2118, Jul. 1999.

[51] J. Serra, *Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances*. London, U.K.: Academic, 1988.

[52] J. Serra, *Image Analysis and Mathematical Morphology*. London, U.K.: Academic, 1982.

[53] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2042–2055, Sep. 2002.

[54] P. Soille, "Recent developments in morphological image processing for remote sensing," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 7477, pp. 747702-1–747702-11, 2009.

[55] J. Crespo, J. Serra, and R. W. Schafer, "Theoretical aspects of morphological filters by reconstruction," *Signal Process.*, vol. 47, no. 2, pp. 201–225, 1995.

[56] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[57] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[58] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C.-W. Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2803–2813, Oct. 2008.

[59] W. Liao, R. Bellens, A. Pizurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 5, no. 4, pp. 1177–1190, 2012

[60] E. Aptoula and S. Lefèvre, "A comparative study on multivariate mathematical morphology," *Pattern Recognit.*, vol. 40, no. 11, pp. 2914–2929, 2007.

[61] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles," *Pattern Recognit.*, vol. 37, no. 6, pp. 1097–1116, 2004.

[62] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed. New York: Springer-Verlag, 2006.

[63] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2005, vol. 1, pp. 176–179.

[64] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–14, Jan. 2009.

[65] T. Castaing, B. Waske, J. A. Benediktsson, and J. Chanussot, "On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5921–5939, 2010.

[66] P. Soille, "Beyond self-duality in morphological image analysis," *Image Vis. Comput.*, vol. 23, no. 2, pp. 249–257, 2005.

[67] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.

[68] J. Chanussot, J. A. Benediktsson, and M. Pesaresi, "On the use of morphological alternated sequential filters for the classification of remote sensing images from urban areas," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2003, pp. 473–475.

[69] J. Debayle and J.-C. Pinoli, "General adaptive neighborhood image processing—Part I," *J. Math. Imag. Vis.*, vol. 25, no. 2, pp. 245–266, 2006.

[70] J. Debayle and J.-C. Pinoli, "General adaptive neighborhood image processing—Part II," *J. Math. Imag. Vis.*, vol. 25, no. 2, pp. 267–284, 2006.

[71] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, Apr. 1990.

[72] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York: Springer-Verlag, 1999.

[73] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[74] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, *Imaging Spectroscopy Special Issue*, no. Suppl. 1, pp. S110–S122, 2009.

[75] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[76] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for support vector machines," *Advances in Neural Information Processing Systems 13*. Cambridge, U.K.: MIT Press, pp. 668–674, 2001.

[77] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[78] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 1884–1887.

[79] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations*. New York: Springer-Verlag, 2001.

[80] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.

[81] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[82] M. Dalla Mura, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "The evolution of the morphological profile: From panchromatic to hyperspectral images," in *Optical Remote Sensing*, vol. 3, S. Prasad, L. M. Bruce, J. Chanussot, R. I. Hammoud, and L. B. Wolff, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 123–146.

[83] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[84] S. Valero, Ph. Salembier, and J. Chanussot, "Hyperspectral image segmentation using binary partition trees," in *Proc. IEEE Int. Conf. Image Process.*, 2011, pp. 1273–1276.

[85] D. Brunner and P. Soille, "Iterative area filtering of multichannel images," *Image Vis. Comput.*, vol. 25, no. 8, pp. 1352–1364, Aug. 2007.

[86] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[87] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 4 , no. 1, pp. 65–74, 2011.

[88] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recognit.*, vol. 13, no. 1, pp. 3–16, 1981.

[89] J. C. Tilton, "Image segmentation by region growing and spectral clustering with a natural convergence criterion," in *Proc. Int. Geosci. Remote Sens. Symp.*, 1998, vol. 4, pp. 1766–1768.

[90] G. Celeux and G. Govaert, "A classification EM algorithm for clustering and two stochastic versions," *Comput. Stat. Data Anal.*, vol. 14, no. 3, pp. 315–332, Oct. 1992.

[91] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised segmentation of satellite images," *IEEE Trans. Geos. Remote Sens.*, vol. 31, no. 3, pp. 618–633, May 1993.

[92] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," presented at the Int. Workshop Image Process. Real-Time Edge Motion Detection/ Estimation, Rennes, France, Sep. 1979.

[93] G. Noyel, J. Angulo, and D. Jeulin, "Morphological segmentation of hyperspectral images," *Image Anal. Stereol.*, vol. 26, pp. 101–109, 2007.

[94] A. N. Evans and X. U. Liu, "A morphological gradient approach to color edge detection," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1454–1463, Jun. 2006.

[95] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[96] L. Shapiro and G. Stockman, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[97] A. C. Jensen and A. S. Solberg, "Fast hyperspectral feature reduction using piecewise 88 constant function approximations," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 547–551, Oct. 2007.

[98] J.-M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: A stepwise optimization approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 2, pp. 150–163, Feb. 1989.

[99] J. C. Tilton, *RHSeg User's Manual: Including HSWO, HSeg, HSegExtract, HSegReader and HSegViewer*, Ver. 1.55, Jan. 2012.

[100] A. J. Plaza and J. C. Tilton, "Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images," in *Proc. Int. Geosci. Remote Sens. Symp.*, Jul. 2005, vol. 7, pp. 4946–4949.

[101] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "Marker-based hierarchical segmentation and classification approach for hyperspectral imagery," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, May 2011, pp. 1089–1092.

[102] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 5, no. 1, pp. 262–272, Feb. 2012.

[103] A. Widayati, B. Verbist, and A. Meijerink, "Application of combined pixel-based and spatial-based approaches for improved mixed vegetation classification using IKONOS," in *Proc. 23rd Asian Conf. Remote Sens.*, pp. 1–8, 2002.

[104] S. V. d. Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert, "Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines," *J. Appl. Remote Sens.*, vol. 1, no. 013543, pp. 1–17, 2007.

[105] C. Chang and C. Lin, *LIBSVM—A Library for Support Vector Machines*, 2008. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[106] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, no. 5, pp. 975–1005, 2004.

[107] G. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geos. Remote Sens.*, vol. 40, no. 10, pp. 2291–2299, Oct. 2002.

[108] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[109] J. Stawiaski, "Mathematical morphology and graphs: Application to interactive medical image segmentation," Ph.D. dissertation, Ctr. Math. Morphology, Paris School of Mines, Paris, France, 2008.

[110] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Technol. J.*, vol. 36, pp. 1389–1401, 1957.

[111] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2008–2021, Apr. 2012.

[112] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.

[113] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 388–400, Apr. 1993.

[114] B. C. Kuo and D. A. Landgrebe, "A robust classification procedure based on mixture classifiers and nonparametric weighted feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2486–2494, Nov. 2002.

ABOUT THE AUTHORS

Mathieu Fauvel received the M.S. and Ph.D. degrees in image and signal processing from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 2004 and 2007, respectively.

In 2007, he was a Teaching Assistant at Grenoble INP. From 2008 to 2010, he was a Postdoctoral Research Associate with the MISTIS Team of the National Institute for Research in Computer Science and Control (INRIA). Since 2010, he has been an Assistant Professor with the National Polytechnic Institute of Toulouse (ENSAT-University of Toulouse) within the DYNAFOR lab (University of Toulouse-INRA), Castanet-Tolosan, France. His research interests are remote sensing, data fusion, pattern recognition, multicomponent signal, and image processing.

Yuliya Tarabalka (Member, IEEE) received the B.S. degree in computer science from Ternopil Ivan Pul'uj State Technical University, Ternopil, Ukraine, in 2005, the M.Sc. degree in signal and image processing from the Grenoble Institute of Technology (INPG), Grenoble, France, in 2007, the Ph.D. degree in signal and image processing from INPG, in 2010, and the Ph.D. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 2010.

From July 2007 to January 2008, she was a Researcher with the Norwegian Defence Research Establishment, Norway. From September 2010 to December 2011, she was a Postdoctoral Research Fellow with CISTO, NASA Goddard Space Flight Center, Greenbelt, MD. Currently, she is a

Postdoctoral Researcher Fellow with INRIA Sophia Antipolis-Méditerranée (AYIN team) and with the French Space Agency (CNES), Paris, France. Her research interests are in the areas of image processing, pattern recognition, hyperspectral imaging, and development of efficient algorithms.

Jón Atli Benediktsson (Fellow, IEEE) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

Currently, he is a Pro Rector for Academic Affairs and Professor of Electrical and Computer Engineering at the University of Iceland. He is a cofounder of the biomedical startup company Oxymap. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields.

Prof. Benediktsson is the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 1999. He was Editor of the IEEE Transactions on Geoscience and Remote Sensing (TGRS) from 2003 to 2008. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award; in 2000, he was granted the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award; in 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland; in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society; and in 2012, he was the corecipient of the IEEE Transactions on Geoscience and Remote Sensing Best Paper Award. He is a member of Societas Scinetiarum Islandica and Tau Beta Pi.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995 and the Ph.D. degree in electrical engineering from Savoie University, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (DGA-French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of Signal and Image Processing. He is currently conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot is the founding President of the IEEE Geoscience and Remote Sensing Society French chapter (2007–2010), which received the 2010 IEEE GRS-S Chapter Excellence Award "for excellence as a Geoscience and Remote Sensing Society chapter demonstrated by exemplary activities during 2009." He was the recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 Symposium Best Paper Award, and the IEEE GRSS 2012 Transactions Prize Paper Award. He was a member of the IEEE Geoscience and Remote Sensing Society AdCom (2009–2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS). He is the Chair (2009–2011) and was the Cochair of the GRS Data Fusion Technical Committee (2005–2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He was an Associate Editor for the IEEE Geoscience and Remote Sensing Letters (2005–2007) and for *Pattern Recognition* (2006–2008). Since 2007, he has been an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing. Since 2011, he has been the Editor-in-Chief of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

James C. Tilton (Senior Member, IEEE) received the B.A. degrees in electronic engineering, environmental science and engineering, and anthropology, the M.E.E. degree in electrical engineering from Rice University, Houston, TX, in 1976, the M.S. degree in optical sciences from the University of Arizona, Tucson, in 1978, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1981.

Currently, he is a Computer Engineer with the Computational and Information Sciences and Technology Office (CISTO) of the Science and Exploration Directorate at the NASA Goddard Space Flight Center, Greenbelt, MD. As a member of CISTO, he is responsible for designing and developing computer software tools for space and Earth science image analysis, and encouraging the use of these computer tools through interactions with space and Earth scientists. His software development has resulted in two patents and two other patent applications.

Dr. Tilton is a Senior Member of the IEEE Geoscience and Remote Sensing Society (GRSS). From 1992 through 1996, he served as a Member of the IEEE GRSS Administrative Committee. Since 1996, he has served as an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing.